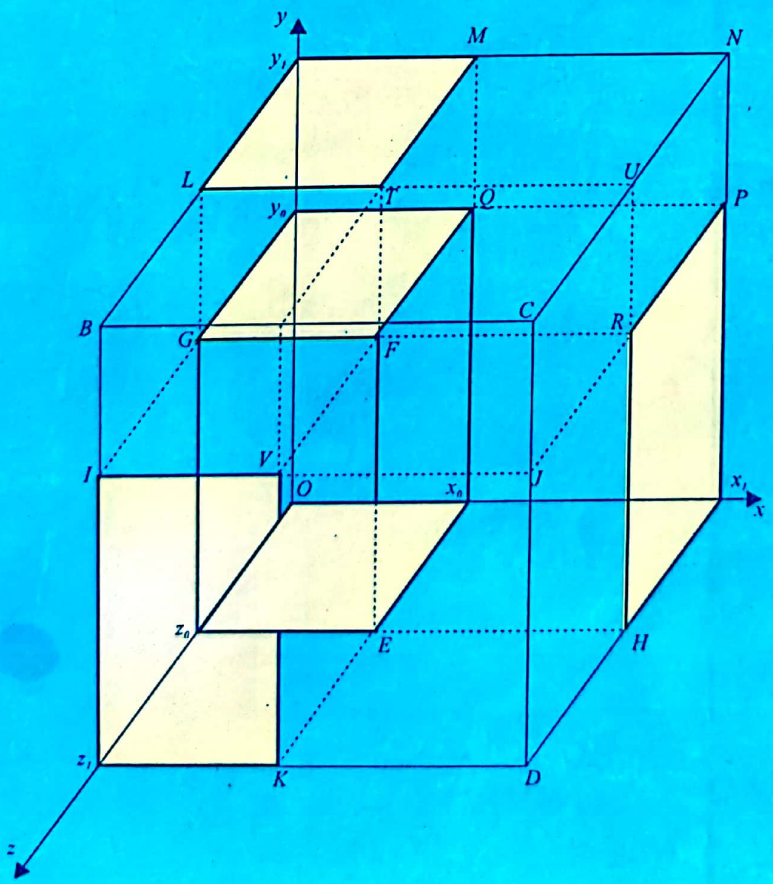
MIHAI ȚARCĂ

TRATAT de STATISTICĂ APLICATĂ



EDITURA DIDACTICĂ ȘI PEDAGOGICĂ, R.A.
BUCUREȘTI

Scanned with OKEN Scanner

Prof. univ. dr. MIHAI ŢARCĂ
Universitatea „Alexandru Ioan Cuza" din Iași

- Statistică
- Statistică matematică.

TRATAT de STATISTICĂ APLICATĂ

Scanned with OKEN Scanner

© Toate drepturile acestei ediţii sunt rezervate Editurii Didactice şi Pedagogice, R.A. Bucureşti.

ISBN 973-30-5275-2

Scanned with OKEN Scanner

Cuvânt înainte

Sfârşitul secolului XX va marca împlinirea a 140 de ani de la înfiinţarea primelor organe ale statisticii de stat din Moldova şi Muntenia, ca şi de la apariţia primului manual de statistică din România care purta modestul titlu de "Elemente de statistică" şi constituia prima parte a lucrării lui Ion Ionescu de la Brad "Povăţuiri pentru facerea catagrafiei Moldovei precedate de oarecari elemente de statistică".

Tratatul de Statistică, *pe care îl punem cu acest prilej la dispoziţia studenţilor economişti, publicat într-o primă variantă în urmă cu aproape două decenii, este conceput ca o lucrare de sinteză a noţiunilor, principiilor şi metodelor statistice, oferind astfel o imagine cuprinzătoare atât asupra stadiului atins de această ştiinţă în prezent, cât şi asupra posibilităţilor şi limitelor în care pot fi folosite metodele statistice pentru analiza şi prognoza dezvoltării fenomenelor social-economice. Elaborarea acestei lucrări s-a făcut pe baza consultării de către autor a unei bogate literaturi de specialitate din ţară şi din străinătate.*

Fiind o lucrare cu caracter metodologic, tratatul de statistică îmbină în mod armonios prezentarea noţiunilor teoretice cu necesităţile practicii economice şi sociale. Expunerea teoretică este completată cu aplicaţii practice bazate pe date reale din statistica românească şi din cea internaţională, iar rezultatele sunt analizate şi uneori ilustrate grafic. Acest mod de tratare sperăm să fie de un real folos cititorilor din domenii variate ale ştiinţelor sociale ca şi celor care studiază fenomenele naturii.

Lucrarea este structurată pe 25 de capitole grupate în şapte secţiuni:

I. Probleme ale Statisticii ca ştiinţă;
II. Obţinerea, sistematizarea şi prezentarea informaţiilor statistice;
III. Analiza seriilor de repartiţie;
IV. Teoria şi practica sondajelor;
V. Analiza legăturilor dintre fenomene;
VI. Analiza seriilor dinamice;
VII. Indicii.
Anexe
Bibliografie.

Scanned with OKEN Scanner

Între lucrările statistice şi cele de matematică publicate până în prezent în ţara noastră există multe dezacorduri, începând cu cele de ordin formal (notaţii diferite pentru elemente similare) şi până la cele de ordin conceptual (interpretarea în termeni de certitudine a unor evenimente probabile etc.).

Având în vedere faptul că metodologia statistică se fundamentează pe teoria domeniului de aplicare, pe teoria probabilităţilor şi statistica matematică, am încercat, prin maniera de tratare a problemelor în această lucrare să înlătur neconcordanţele mai sus semnalate.

Prin problematica vastă pe care o abordează, tratatul nu numai că acoperă în întregime programele analitice în vigoare pentru studenţii economişti, ci prezintă o dezvoltare mai amplă a problemelor, fiind astfel util şi doctoranzilor, participanţilor la cursurile postuniversitare, cercetătorilor şi cadrelor didactice din învăţământul superior, precum şi practicienilor din diferite domenii ale activităţii sociale şi economice.

Expunerea metodelor statistice se limitează la noţiunile de matematică strict necesare pentru justa înţelegere a metodelor statistice, evitându-se peste tot unde a fost posibil, utilizarea unui aparat matematic cu înalt grad de abstractizare.

Având în vedere această orientare, lucrarea poate constitui un îndreptar preţios şi o continuare în direcţia aplicaţiilor practice atât pentru lucrările teoretice de statistică matematică, informatică şi econometrie, cât şi pentru lucrările aplicative de economie, demografie, antropologie, sociologie, biologie etc. în vederea cuantificării, prelucrării ştiinţifice, analizei şi prognozei fenomenelor din natură şi societate pe baza informaţiilor existente în aceste domenii.

Autorul este conştient de faptul că şi acestă lucrare este susceptibilă de îmbunătăţiri şi va fi recunoscător tuturor cititorilor care vor binevoi a-i face observaţii, sugestii şi propuneri care să ducă la ridicarea nivelului ei calitativ într-o eventuală nouă ediţie.

Iaşi
1 Decembrie 1997

Prof. univ. dr. Mihai Ţarcă
Universitatea "Al.I.Cuza" Iaşi
Facultatea de Ştiinţe Economice

Scanned with OKEN Scanner

CUPRINS

Cuvânt înainte
Cuprins 1
Table des matières 11
Contents 21

I. PROBLEME ALE STATISTICII CA STIINTĂ 31

1. INTRODUCERE 33

1.1. Revoluţia ştiinţifică contemporană şi statistica 33
1.2. Din istoria statisticii 35
1.3. Cunoaşterea statistică a realităţii 39
1.4. Legile statistice 43
1.5. Metoda particulară a statisticii 44
1.6. Definiţia statisticii 48

2. NOŢIUNI FUNDAMENTALE ALE STATISTICII 51

2.1. Colectivităţi statistice 51
2.2. Unităţi statistice 54
2.3. Caracteristicile unităţilor statistice 55
2.4. Cuantificarea fenomenelor sociale 60

II. OBTINEREA, SISTEMATIZAREA ŞI PREZENTAREA INFORMAŢIILOR STATISTICE 63

3. OBŢINEREA INFORMAŢIILOR STATISTICE 65

3.1. Înregistrarea statistică şi felurile ei 66
3.2. Verificarea autenticităţii informaţiilor statistice 69
3.2.1. Clasificarea erorilor de înregistrare 70
3.2.2. Prevenirea şi controlul statistic al erorilor 71

Scanned with OKEN Scanner

4. GRUPAREA INFORMAŢIILOR STATISTICE 73

4.1. Noţiunea de grupare a unităţilor statistice 73
4.2. Felurile grupărilor 74
4.3. Stabilirea intervalelor de grupare 78

5. PREZENTAREA INFORMAŢIEI STATISTICE 83

5.1. Serii de repartiţie unidimensionale 83
5.1.1. Noţiunea, importanţa şi felurile lor 83
5.1.2. Reprezentarea grafică a seriilor de repartiţie 92
5.2. Serii de repartiţie multidimensionale 106
5.2.1. Serii de repartiţie bidimensionale 106
5.2.1.1. Noţiunea, importanţa şi felurile lor 106
5.2.1.2. Reprezentarea grafică a seriilor de distribuţie bidimensionale
5.2.2. Distribuţii tridimensionale 113
5.3. Seriile dinamice 119
5.3.1. Noţiunea, importanţa şi felurile lor 119
5.3.2. Reprezentarea grafică a seriilor dinamice 124
5.3.2.1. Historiogramele 125
5.3.2.2. Diagramele polare 131
5.4. Seriile de spaţiu 135
5.4.1. Noţiunea şi importanţa lor 135
5.4.2. Reprezentarea grafică a seriilor de spaţiu 135

III. ANALIZA DISTRIBUTIILOR UNIDIMENSIONALE 141

6. MĂRIMILE MEDII 143

6.1. Esenţa şi rolul valorilor medii în cunoaşterea statistică 143
6.2. Medii propriu-zise 147
6.2.1. Media aritmetică 147
6.2.1.1. Proprietăţile mediei aritmetice 151
6.2.1.2. Avantajele şi dezavantajele mediei aritmetice 167
6.2.1.3. Media aritmetică a unei caracteristici alternative 168
6.2.2. Media armonică 169
6.2.2.1. Proprietăţile mediei armonice 171

Scanned with OKEN Scanner

6.2.2.2. Determinarea mediei armonice pe cale grafică 175
6.2.3. Media anarmonică 176
6.2.4. Media geometrică 177
6.2.4.1. Proprietăţile mediei geometrice 179
6.2.5. Media pătratică 186
6.2.5.1. Proprietăţile mediei pătratice 187
6.2.6. Generalizarea noţiunii de medie 187
6.3. Medii de poziţie 193
6.3.1. Modul 193
6.3.1.1. Determinarea modului în cazul variabilelor discrete 194
6.3.1.2. Determinarea modului în cazul unei variabile continue 195
6.3.1.2.1. Determinarea modului în cazul unor serii cu intervale egale
6.3.1.2.1.1. Determinarea grafică a modului 200
6.3.1.2.1.2. Stabilirea formulei de calcul al modului în cazul unor serii cu intervale egale 202
6.3.1.2.1.3. Determinarea modului în cazul unor serii cu intervale neegale
6.3.1.3. Proprietăţile modului 203
6.3.2. Mediana 206
6.3.2.1. Determinarea grafică a medianei pentru o serie cu intervale 212
6.3.2.2. Proprietăţile medianei 213
6.3.2.3. Cuantile 214
6.3.3. Mediala 217

7. INDICATORII VARIAŢIEI ŞI CONCENTRĂRII 223

7.1. Indicatorii simpli ai variaţiei 224
7.2. Indicatorii sintetici ai variaţiei 226
7.2.1. Proprietăţile dispersiei şi calculul ei simplificat 236
7.2.2. Indicatorii factoriali ai dispersiei 253
7.3. Indicatorii variaţiei pentru o caracteristică alternativă 256
7.4. Concentrare 257
7.4.1 Proprietăţile curbei de concentrare.....264
7.4.2 Indicele de concentrare.....266

8. REPARTIŢIILE UNOR VARIABILE ALEATOARE 277

8.1. Repartiţia binomială 277
8.2. Repartiţia Poisson 283
8.3. Repartiţia normală 285
8.3.1. Proprietăţile distribuţiei normale 287

Scanned with OKEN Scanner

9. CARACTERISTICI ALE FORMEI DISTRIBUŢIILOR.295
9.1. Asimetria distribuţiilor.....295
9.2. Boltirea distribuţiilor.....305

IV. TEORIA SI PRACTICA SONDAJELOR.....307

10. SONDAJELE STATISTICE.....309

10.1. Noţiunea, importanţa şi felurile sondajelor.....309
10.1.1. Noţiuni fundamentale.....309
10.1.2. Reprezentativitatea eşantionului.....312
10.1.3. Avantajele sondajelor.....314
10.1.4. Tipuri de sondaje.....315
10.2. Sondaje de volum mare.....318
10.2.1. Sondaj întâmplător simplu.....318
10.2.1.1. Sondajul întâmplător simplu repetat.....318
10.2.1.2. Sondajul întâmplător simplu nerepetat.....335
10.2.1.3. Sondajul mecanic.....337
10.2.1.4. Dimensionarea eşantioanelor întâmplătoare.....338
10.2.2. Sondajul de serii.....340
10.2.3. Sondajul tipic.....344
10.2.3.1. Sondajul tipic simplu.....347
10.2.3.2. Sondajul tipic proporţional.....348
10.2.3.3. Sondajul tipic optim.....349

11. VERIFICAREA IPOTEZELOR STATISTICE.....353

11.1. Noţiunea de estimare.....353
11.2. Verificarea ipotezelor.....355
11.2.1. Compararea mediei unei colectivităţi cu o valoare dată.....357
11.2.2. Compararea mediilor a două eşantioane.....361
11.2.3. Testarea erorii standard a medianei, cuartilelor şi decilelor...364
11.2.4. Eroarea standard a abaterii medii pătratice şi a coeficientului de variaţie.....366

Scanned with OKEN Scanner

V. ANALIZA LEGĂTURILOR DINTRE FENOMENE.....371

12. ANALIZA DISPERSIONALĂ.....373

12.1. ANOVA unifactorială.....374
12.1.1. ANOVA unifactorială cu eşantioane de volum *n*.....374
12.1.2. Analiza dispersională unifactorială pe baza unor eşantioane de volum diferit.....382
12.2. ANOVA bifactorială.....386
12.2.1. ANOVA bifactorială cu o singură observaţie în fiecare grup.386
12.2.2. ANOVA bifactorială cu eşantioane de acelaşi volum.....388

13. LEGĂTURILE DINTRE FENOMENELE REALE.....395

13.1. Noţiuni preliminare.....395
13.2. Metode elementare de studiere a legăturilor dintre fenomene..398
13.2.1. Metoda seriilor statistice interdependente.....398
13.2.2. Metoda grupărilor.....399
13.2.3. Metoda grafică.....400

14. REGRESIE ŞI CORELAŢIE SIMPLĂ LINIARĂ.....403

14.1. Regresie simplă liniară.....403
14.1.1. Metode de rezolvare a sistemelor de ecuaţii normale liniare 408
14.2. Dualitate în regresie.....429
14.3. Intensitatea legăturilor liniare.....433
14.3.1. Coeficientul de corelaţie.....433
14.3.2. Raportul de corelaţie.....445
14.3.3. Legătura dintre coeficienţii de regresie şi coeficientul de corelaţie.....457

15. REGRESIE ŞI CORELAŢIE SIMPLĂ CURBILINIE..461

15.1. Modele de regresie simplă curbilinie.....461
15.1.1. Legături de tip parabolic.....462
15.1.2. Legături de tip hiperbolic.....476
15.1.3. Legături după modelul unei funcţii putere.....491
15.1.4. Legături de tip exponenţial.....496

Scanned with OKEN Scanner

15.1.5. Legături de tip logaritmic.....504
15.1.6. Alegerea celei mai potrivite ecuaţii de regresie.....509

16. REGRESIE ŞI CORELAŢIE MULTIPLĂ LINIARĂ...513

16.1. Noţiunea de legătură multiplă.....513
16.2. Regresie şi corelaţie liniară multiplă.....514
16.2.1. Regresie liniară multiplă.....514
16.2.2. Intensitatea corelaţiei multiple liniare.....518
16.3. Corelaţie parţială.....522
16.4. Determinaţie multiplă şi parţială.....524

17. REGRESIE ŞI CORELAŢIE MULTIPLĂ CURBILINIE533

17.1. Regresie multiplă curbilinie.....533
17.1.1. Liniarizarea legăturilor multiple curbilinii.....536
17.2. Intensitatea corelaţiei multiple curbilinii.....538
17.2.1. Determinaţie curbilinie multiplă şi parţială.....540

18. CORELAŢIE NEPARAMETRICĂ.....547

18.1. Noţiunea de corelaţie neparametrică.....547
18.2. Coeficientul de corelaţie a rangurilor al lui Spearman.....548
18.3. Coeficientul de corelaţie a rangurilor al lui Kendall.....553
18.4. Coeficientul de concordanţă al lui Fechner.....557

19. ASOCIAŢIA ÎNTRE CARACTERISTICI CALITATIVE

19.1. Noţiunea de asociaţie.....561
19.2. Existenţa asociaţiei.....562
19.3. Intensitatea asociaţiei.....570

20. ELASTICITATEA.....575

20.1. Noţiunea şi felurile ei.....575
20.2. Elasticitate simplă.....576
20.3. Elasticitate multiplă.....582

Scanned with OKEN Scanner

21. FUNCŢIILE DE PRODUCŢIE.....583

21.1. Noţiunea şi felurile lor.....583
21.2. Funcţia putere cu un singur factor.....585
21.3. Funcţiile putere cu doi factori.....587
21.3.1. Reprezentarea grafică a funcţiilor putere cu doi factori.....589
21.3.2. Proprietăţile funcţiilor de producţie.....591
21.3.3. Produs mediu şi produs marginal.....593
21.3.4. Estimarea parametrilor unei funcţii putere cu doi factori.....596
21.4. Funcţia putere cu trei factori.....598

VI. ANALIZA VARIATIEI FENOMENELOR IN TIMP.....601

22. INDICATORII SERIILOR DINAMICE.....603

22.1. Indicatorii nivelului absolut şi cei ai sporului.....604
22.2. Indicele variaţiei totale a fenomenelor.....611
22.3. Indicele creşterii fenomenelor.....612
22.3.1. Indicele mediu de creştere.....614
22.3.1.1. Metoda mediei geometrice.....615
22.3.1.2. Metoda mediei parabolice.....617
22.3.1.3. Metoda trendului exponenţial.....618
22.3.2. Determinarea indicelui mediu pentru fenomene complexe...620
22.4. Prognoza fenomenelor pe baza indicatorilor medii de creştere625
22.4.1. Metode de prognoză bazate pe sporul mediu.....625
22.4.1.1. Determinarea nivelului unui fenomen peste t ani.....626
22.4.1.2. Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit.....626
22.4.1.3. Stabilirea perioadei de timp t necesară pentru egalizarea nivelului de dezvoltare a două zone A şi B.....627
22.4.1.4. Egalizarea nivelului de dezvoltare a două zone peste o perioadă de timp prestabilită.....628
22.4.2. Metode de prognoză bazate pe indicele mediu al variaţiei....628
22.4.2.1. Determinarea nivelului unui fenomen peste t ani.....629
22.4.2.2. Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit.....629
22.4.2.3 Stabilirea perioadei de timp t necesară pentru egalizarea nivelului de dezvoltare a două zone A şi B.....630

Scanned with OKEN Scanner

22.4.2.4. Egalizarea nivelului de dezvoltare a două zone peste o perioadă de timp prestabilită.....630

23. TENDINŢA GENERALĂ DE DEZVOLTARE A FENOMENELOR ÎN TIMP.....633

23.1. Componentele principale ale evoluţiei fenomenelor în timp....633
23.2. Metode de ajustare a seriilor dinamice.....637
23.2.1. Metoda grafică.....638
23.2.2. Metoda mediilor mobile.....639
23.2.3. Metoda mediilor înjumătăţite.....642
23.2.4. Metoda sporului mediu.....644
23.2.5. Metoda indicelui mediu de variaţie.....647
23.2.6. Metoda celor mai mici pătrate.....649
23.2.6.1. Tendinţa liniară.....650
23.2.6.2. Tendinţa parabolică de gradul doi.....655
23.2.6.3. Tendinţa parabolică de gradul trei.....660
23.2.6.4. Tendinţa exponenţială.....664
23.2.7. Ajustarea după o curbă exponenţială modificată.....667
23.2.8. Curba lui Gompertz.....677
23.2.9. Curba logistică.....683
23.2.10. Curba exponenţială de gradul doi.....691
23.3. Interpolarea seriilor dinamice.....693
23.3.1.Formula de interpolare a lui Lagrange.....694
23.3.2. Formula de interpolare a lui Newton.....695

24. STUDIUL VARIAŢIILOR SEZONIERE, CICLICE ŞI A CELOR ALEATOARE.....697

24.1. Variaţiile sezoniere.....697
24.2. Măsurarea variaţiilor sezoniere.....702
24.2.1. Analiza sezonalităţii în cazul modelului multiplicativ.....702
24.2.2. Analiza sezonalităţii în cazul modelului aditiv.....711
24.3. Variaţiile ciclice.....715
24.4. Variaţiile aleatoare.....717

Scanned with OKEN Scanner

VII. INDICII.....719

25. INDICI.....721

25.1. Noţiunea şi importanţa indicilor.....721
25.2. Clasificarea indicilor.....724
25.3. Indicii dinamicii.....725
25.3.1. Indicii simpli sau elementari.....726
25.3.2. Indicii de grup sau sintetici.....730
25.3.2.1. Indicii agregaţi.....730
25.3.2.2. Indicele mediu aritmetic ponderat.....746
25.3.2.3. Indicele mediu armonic ponderat.....747
25.3.2.4. Indicele mediu geometric.....748
25.3.2.5. Indici calculaţi ca raport între două medii aritmetice ponderate.....750
25.4. Analiza influenţei factorilor cu ajutorul indicilor.....757
25.4.1. Cu ajutorul indicilor agregaţi.....757
25.4.2. Cu ajutorul indicilor valorilor medii.....762
25.5. Scheme de descompunere a indicilor pe factori.....767
25.5.1. Schema aditivă.....767
25.5.2. Schema multiplicativă.....769
25.5.2.1. Metoda influenţelor separate.....769
25.5.2.2. Metoda substituirii în lanţ.....782
25.6. Indicii teritoriali.....788
25.7. Indicii de plan.....793

ANEXE.....795
BIBLIOGRAFIE.....797

Scanned with OKEN Scanner

Table des matières

Cuprins.....1
Table des matières.....11
Contents.....21

I.PROBLEMES DE LA STATISTIQUE EN TANT QUE SCIENCE.....31

1. INTRODUCTION.....33

1.1. La révolution scientifique contemporaine et la statistique.....33
1.2. De l'histoire de la statistique.....35
1.3. La connaissance statistique de la réalité.....39
1.4. Les lois statistiques.....43
1.5. La méthode particulière de la statistique.....44
1.6. Définition de la statistique.....48

2. NOTIONS FONDAMENTALES DE LA STATISTIQUE.....51

2.1. Collectivités statistiques.....51
2.2. Unités statistiques.....54
2.3. Les caractéristiques des unités statistiques.....55
2.4. La quantification des fénomènes sociaux.....60

II. LE GROUPEMENT, LA SYSTEMATISATION ET LA PRESENTATION DES INFORMATIONS STATISTIQUES.....63

3. OBTENTION DES INFORMATIONS STATISTIQUES.....65

3.1. L'enregistrement statistique et ses types.....66
3.2. Vérification de l'authenticité des informations statistiques.....69
3.2.1. Classification des erreurs d'enregistrement.....70

Scanned with OKEN Scanner

3.2.2. Prévention et contrôle statistique des erreurs.....71

4. LE GROUPEMENT DES INFORMATIONS STATISTIQUES.....73
4.1. La notion de groupement des unités statistiques.....73
4.2. Types de groupements.....74
4.3. Etablir les intervales de groupement.....78

5. LA PRESENTATION DE L'INFORMATION STATISTIQUE.....83

5.1. Séries de répartition unidimensionnelles.....83
5.1.1. La notion, son importance et ses types.....83
5.1.2. Représentation graphique des séries de répartition.....92
5.2. Séries de répartition multidimensionnelle.....106
5.2.1. Séries de répartition bidimensionnelles.....106
5.2.1.1. La notion, son importance et ses types.....106
5.2.1.2. La représentation graphique des séries de distribution bidimensionnelles.....110
5.2.2. Distributions tridimensionnelles.....113
5.3. Séries dynamiques.....119
5.3.1. La notion, son importance et ses types.....119
5.3.2. Représentation graphique des séries dynamiques.....124
5.3.2.1. Les historiogrammes.....125
5.3.2.2. Les diagrammes polaires.....131
5.4. Les séries d'espace.....135
5.4.1. La notion et son importance.....135
5.4.2. La représentation graphique des séries d'espace.....135

III. ANALYSE DES DISTRIBUTIONS UNIDIMENSIONNELLES.....141

6. LES MOYENNES.....143

6.1. Essence et rôle des moyennes dans la connaissance statistique.....143
6.2. Les moyennes proprement-dites.....147
6.2.1. La moyenne arithmétique.....147
6.2.1.1. Propriétés de la moyenne arithmétique.....151
6.2.1.2. Avantages et désavantages de la moyenne arithmétique.....167
6.2.1.3. La moyenne arithmétique d'une caractéristique alternative.....168

Scanned with OKEN Scanner

6.2.2. La moyenne harmonique.....169
6.2.2.1. Propriétés de la moyenne harmonique.....171
6.2.2.2. Déterminer la moyenne harmonique à l'aide des graphiques.....175
6.2.3. La moyenne anharmonique.....176
6.2.4. La moyenne géométrique.....177
6.2.4.1. Propriétés de la moyenne géométrique.....179
6.2.5. La moyenne quadratique.....186
6.2.5.1. Propriétés de la moyenne quadratique.....187
6.2.6. Généralisation de la notion de moyenne.....187
6.3. Moyennes de position.....193
6.3.1. Le mode.....193
6.3.1.1. Détermination du mode dans le cas des variables discrètes.....194
6.3.1.2. Détermination du mode dans le cas d'une variable continue.....195
6.3.1.2.1. Détermination du mode dans le cas des séries à intervales égaux.....196
6.3.1.2.1.1. Détermination graphique du mode.....200
6.3.1.2.1.2. Détermination de la formule de calcul du mode dans le cas des séries à intervales égaux.....203
6.3.1.2.1.3. Détermination du mode dans le cas des series à intervales inégaux.....204
6.3.1.3. Les propriétés du mode.....204
6.3.2. La médiane.....206
6.3.2.1. La détermination graphique de la médiane pour une série avec intervales.....212
6.3.2.2. Les propriétés de la médiane.....213
6.3.2.3. Quantiles.....214
6.3.3. La médiale.....217

7. INDICATEURS DE LA VARIATION ET DE LA CONCENTRATION.....223

7.1. Indicateurs simples de la variation.....224
7.2. Indicateurs synthétiques de la variation.....226
7.2.1. Les propriétés de la dispersion et son calcul simplifié.....236
7.2.2. Les indicateurs factoriels de la dispersion.....253
7.3. Les indicateurs de la variation pour une caractéristique alternative.256
7.4. La concentration.....257
7.4.1. Les propriétés de la courbe de concentration.....264

Scanned with OKEN Scanner

7.4.2. L'indice de concentration.....266

8. LA REPARTITION DE CERTAINES VARIABLES ALEATOIRES.....277

8.1. La répartition binomiale.....277
8.2. La répartition Poisson.....283
8.3. La répartition normale.....285
8.3.1. Les propriétés de la distribution normale.....287

9. CARACTERISTIQUES DE LA FORME DES DISTRIBUTIONS.295

9.1. Asymétrie des distributions.....295
9.2. Voussure des distributions.....305

IV. THEORIE ET PRATIQUE DES SONDAGES.....307

10. LES SONDAGES STATISTIQUES.....309

10.1. La notion, son importance et les types de sondages.....309
10.1.1. Notions fondamentales.....309
10.1.2. La représentivité de l'échantillon.....312
10.1.3. Les avantages des sondages.....314
10.1.4. Types de sondages.....315
10.2. Les sondages à grand volume.....318
10.2.1. Le sondage aléatoire simple.....318
10.2.1.1. Le sondage aléatoire avec remise.....318
10.2.1.2. Le sondage aléatoire simple sans remise.....335
10.2.1.3. Le sondage mécanique.....337
10.2.1.4. Mesuration des échantillons aléatoire.....338
10.2.2. Le sondage de séries.....340
10.2.3. Le sondage stratifié.....344
10.2.3.1. Le sondage stratifié simple.....347
10.2.3.2. Le sondage stratifié proportionnel.....348
10.2.3.3. Le sondage stratifié optimum.....349

Scanned with OKEN Scanner

11. VERIFICATION DES HYPOTHESES STATISTIQUES.....353

11.1. La notion d'évaluation.....353
11.2. Vérification des hypothèses.....355
11.2.1. Comparer la moyenne d'une collectivité à une valeur donnée....357
11.2.2. Comparer les moyennes de deux échantillons.....361
11.2.3. Tester l'erreur standard de la médiane, des quartiles et des déciles.....364
11.2.4. L'erreur standard de la déviation moyenne quadratique et du coefficient de variation.....366

V. L'ANALYSE DES RELATIONS ENTRE LES PHÉNOMÈNES.....371

12. L'ANALYSE DISPERTIONNELLE.....373

12.1. L'ANOVA unifactorielle.....374
12.1.1. L'ANOVA unifactorielle avec échantillons de volume n.....374
12.1.2. L'analyse dispersionnelle unifactorielle à partir d'échantillons de volume différent.....382
12.2. L'ANOVA bifactorielle.....386
12.2.1. L'ANOVA bifactorielle avec une seule observation dans chaque groupe.....386
12.2.2. L'ANOVA bifactorielle avec échantillons du même volume.....388

13. RELATIONS ENTRE LES PHENOMENES REELS.....395

13.1. Notions préliminaires.....395
13.2. Méthodes élémentaires d'étude des relations entre les phénomènes.....398
13.2.1. La méthode des séries statistiques interdépendentes.....398
13.2.2. La méthode des groupements.....399
13.2.3. La méthode graphique.....400

Scanned with OKEN Scanner

14. REGRESSION ET CORRELATION SIMPLE LINEAIRE.....403

14.1. La regression simple linéaire.....403
14.1.1. Méthodes pour résoudre des systèmes d' écuations normales linéaires.....408
14.2. La dualité en regression.....429
14.3. L'intensité de la liaison linéaire.....433
14.3.1. Le coéficient de correlation.....433
14.3.2. Le rapport de corrélation.....445
14.3.3. Le rapport entre les coéficients de regression et le coéficient de correlation.....457

15. REGRESSION ET CORRELATION SIMPLES CURVILIGNE.....461

15.1. Modèles de regression simple curviligne.....461
15.1.1. Rapports de type parabolique.....462
15.1.2. Rapports de type hyperbolique.....476
15.1.3. Rapports selon le modèle d'une fonction de pouvoir.....491
15.1.4. Rapports de type exponentiel.....496
15.1.5. Rapports de type logarithmique.....504
15.1.6. Le choix de la plus adéquate écuation de regression.....509

16. REGRESSION ET CORRELATION MULTIPLE LINEAIRE.....513

16.1. La notion de liaison multiple.....513
16.2. Regression et correlation linéaire multiple.....514
16.2.1. La regression linéaire multiple.....514
16.2.2. L'intensité de la correlation linéaire multiple.....518
16.3. La correlation partielle.....522
16.4. La détermination multiple et partielle.....524

17. REGRESSION ET CORRELATION MULTIPLE CURVILIGNE.....533

17.1. La regression multiple curviligne.....533
17.1.1. Linéarisation des liaisons multiples curvilignes.....536
17.2. L'intensité de la correlation multiple curviligne.....538

Scanned with OKEN Scanner

17.2.1. La détermination curviligne multiple et partielle.....540

18. LA CORRELATION NON-PARAMÉTRIQUE.....547

18.1. La notion de correlation non-paramétrique.....547
18.2. Le coéficient de correlation des rangs de Spearman.....548
18.3. Le coéficient de correlation de Kendall.....553
18.4. Le coéficient de concordance de Fechner.....557

19. ASSOCIER DES CARACTERISTIQUES QUALITATIVES.....561

19.1. La notion d'association.....561
19.2. L'existence de l'association.....562
19.3. L'intensité de l'association.....570

20. L'ELASTICITE.....575

20.1. La notion et ses types.....575
20.2. L'élasticité simple.....576
20.3. L'élasticité multiple.....582

21. LES FONCTIONS DE PRODUCTION.....583

21.1. La notion et ses types.....583
21.2. La fonction de pouvoir à un seul facteur.....585
21.3. Les fonctions de pouvoir à deux facteurs.....587
21.3.1. La représentation graphique des fonctions de pouvoir à deux facteurs.....589
21.3.2. Les propriétés des fonctions de production.....591
21.3.3. Le produit moyen et le produit marginal.....593
21.3.4. Estimation des paramètres d'une fonction de pouvoir à trois variables.....596
21.4. La fonction de pouvoir à quatre variables.....598

im. 149 foi

Scanned with OKEN Scanner

VI. ANALYSE DE LA VARIATION DES PHENOMENES DANS LE TEMPS.....601

22. LES INDICATEURS DES SERIES DYNAMIQUES.....603

22.1. Les indicateurs du niveau absolu et de l'augmentation 604
22.2. L'indice de la variation totale des phénomènes.....611
22.3. L'indice de l'augmentation des phénomènes.....612
22.3.1. L'indice moyen d'augmentation.....614
22.3.1.1. La méthode de la moyenne géométrique.....615
22.3.1.2. La méthode de la moyenne parabolique.....617
22.3.1.3. La méthode de la tendence exponentielle.....618
22.3.2. Déterminer l'indice moyen pour des phénomènes complexes.....620
22.4. La prognose des phénomènes à partir des indicateurs moyens d'augmentation.....625
22.4.1. Méthodes de prognose à partir de l'augmentation moyenne.....625
22.4.1.1. Déterminer le niveau d'un phénomène après t ans.....626
22.4.1.2. Déterminer l'horizon de la prognose où un certain fenomène atteindra un niveau établi d'avance.....626
22.4.1.3. Déterminer la période de temps t nécessaire pour égaliser le niveau de développement de deux zones A et B.....627
22.4.1.4. L'égalisation du niveau de développement de deux zones après une période de temps établie à l'avance.....628
22.4.2. Méthodes de prognose basées sur l'indice moyen de la variation 628
22.4.2.1. Déterminer le niveau d'un phénomène après t ans.....629
22.4.2.2. Déterminer l'horizon de la prognose où un certain phénomène atteindra un niveau établi à l'avance.....629
22.4.2.3. Déterminer la période de temps t nécessaire pour égaliser le niveau de développement de deux zones A et B.....630
22.4.2.4. Egalisation du niveau de développement de deux zones après une période de temps établie à l'avance.....630

23. LA TENDENCE GÉNÉRALE DE DEVELOPPEMENT DES PHÉNOMÈNES DANS LE TEMPS.....633

23.1. Les principales composantes de l'évolution des phenomenes dans le temps.....633

Scanned with OKEN Scanner

23.2.1. La méthode graphique.....638
23.2.2. La méthode des moyennes mobiles.....639
23.2.3. La méthode des moyennes divisées.....642
23.2.4. La méthode de l'augmentation moyenne.....644
23.2.5. La méthode de l'indice moyen de variation.....647
23.2.6. La méthode des plus petits carrés.....649
23.2.6.1. La tendence linéaire.....650
23.2.6.2. La tendence parabolique de deuxième degré.....655
23.2.6.3. La tendence parabolique de troisième degré.....660
23.2.6.4. La tendence exponentielle.....664
23.2.7. La tendence exponentielle modifiée.....667
23.2.8. La courbe de Gompertz.....677
23.2.9. La courbe logistique.....683
3.2.10. La courbe exponentielle de deuxième degré.....691
23.3. Interpolation des séries dynamiques.....693
23.3.1. La formule d'interpolation de Lagrange.....694
23.3.2. La formule d'interpolation de Newton.....695

24. L'ETUDE DES VARIATIONS SAISONNIERES, CYCLIQUES ET ALEATOIRES.....697

24.1. Les variations saisonnières.....697
24.2. L'évaluation des variations saisonnières.....702
24.2.1. Analyser la saisonnalité dans le cas du modèle multiplicatif.....702
24.2.2. Analyser la saisonnalité dans le cas du modèle aditif.....711
24.3. Les variations cycliques.....715
24.4. Les variations aléatoires.....717

VII. INDICES.....719

25. INDICES.....721

25.1. La notion et l'importance des indices.....721
25.2. Classification des indices.....724
25.3. Les indices de la dynamique.....725
25.3.1. Indices simples ou élémentaires.....726
25.3.2. Indices de groupe ou synthétiques.....730
25.3.2.1. Indices agrégés.....730

Scanned with OKEN Scanner

25.3.2.1. Indices agrégés.....730
25.3.2.2. L'indice moyen arithmétique pondéré.....746
25.3.2.3. L'indice moyen harmonique pondéré.....747
25.3.2.4. L'indice moyen géométrique.....748
25.3.2.5. Indices calculés comme rapport de deux moyennes arithmétiques pondérées.....750
25.4. Analyser l'influence des facteurs à l'aide des indices.....757
25.4.1. A l'aide des indices agrégés.....757
25.4.2. A l'aide des indices des valeurs moyennes.....762
25.5. Schémas de décomposition des indices.....767
25.5.1. Le schéma aditif.....767
25.5.2. Le schéma multiplicatif.....769
25.5.2.1. La méthode des influences séparées.....769
25.5.2.2. La méthode de la substitution en chaîne.....782
25.6. Les indices territoriels.....788
25.7. Les indices de plan.....793

TABLES.....795

BIBLIOGRAPHIE.....797

Scanned with OKEN Scanner

Contents

Cuprins.....1
Table des matières.....11
Contents.....21

I. PROBLEMS CONCERNING THE STATISTICS AS SCIENCE.....31

1. INTRODUCTION.....33

1.1. Statistics and Contemporary Science Development.....33
1.2. History of Statistics.....35
1.3. Statistical Knowledge of Reality.....39
1.4. Statistical Laws.....43
1.5. The Particular Method of Statistics.....44
1.6. Definition of Statistics.....48

2. BASIC NOTIONS OF STATISTICS.....51

2.1. Statistical Population.....51
2.2. Statistical Units.....54
2.3. Characteristics of Statistical Units.....55
2.4. Quantification of Social Phenomena.....60

Scanned with OKEN Scanner

II. REGISTRATION, SYSTEMATIZATION AND PRESENTATION OF STATISTICAL DATA.....63

3. COLLECTING STATISTICAL DATA.....65

3.1. The Statistical Registration and Its Types.....66
3.2. Checking the Authenticity of Statistical Data.....69
3.2.1. Classification of the Registration Errors.....70
3.2.2. Preventing and Controlling Statistical Errors.....71

4. CLASSIFICATION OF STATISTICAL DATA.....73

4.1. Classification of Statistical Units.....73
4.2. Types of Classifications.....74
4.3. Establishing the Intervals of Classification.....78

5. PRESENTATION OF STATISTICAL DATA.....83

5.1. Unidimensional Distribution.....83
5.1.1. Notion, Importance and Types of Distribution.....83
5.1.2. The Graphic Presentation of the Distribution.....92
5.2. Multidimensional Distribution.....106
5.2.1. Bidimensional Distribution.....106
5.2.1.1. Notion, Importance, Types of Bidimensional Distribution....106
5.2.1.2. Graphic Presentation of Bidimensional Distribution.....110
5.2.2. Tridimensional Distributions.....113
5.3. Time Series.....119
5.3.1. Notion, Importance and Types of Time Series.....119
5.3.2. Graphic Presentation of Time Series.....124
5.3.2.1.Historiograms.....125
5.3.2.2. Circular Charts.....131
5.4. Space Distributions.....135
5.4.1. Notion and Importance of Space Distributions.....135
5.4.2. Graphic Presentation of Space Distributions.....135

Scanned with OKEN Scanner

III. ANALYSIS OF UNIDIMENSIONAL DISTRIBUTIONS.....141

6. THE MEANS.....143

6.1. Basic Notions and the Importance of Means in Statistical Knowledge.....143
6.2. Proper Means.....147
6.2.1. The Arithmetic Mean.....147
6.2.1.1. Properties of the Arithmetic Mean.....151
6.2.1.2. Advantages and Disavantages of the Arithmetic Mean.....167
6.2.1.3. The Arithmetic Mean of an Alternative Characteristic.....168
6.2.2. The Harmonic Mean.....169
6.2.2.1. Properties of the Harmonic Mean.....171
6.2.2.2. Establishing the Harmonic Mean by Using the Graphic Method.....175
6.2.3. The Unharmonic Mean.....176
6.2.4. The Geometric Mean.....177
6.2.4.1. Properties of the Geometric Mean.....179
6.2.5. The Quadratic Mean.....186
6.2.5.1. Properties of the Quadratic Mean.....187
6.2.6. Generalization of the Notion of Mean.....187
6.3. Measures of Location.....193
6.3.1. The Mode.....193
6.3.1.1. Establishing the Mode for a Discrete Variable.....194
6.3.1.2. Establishing the Mode for a Continuous Variable.....195
6.3.1.2.1. Establishing the Mode for Equal Interval Series.....196
6.3.1.2.1.1. The Graphic Estimation of the Mode.....200
6.3.1.2.1.2. Establishing the Calculus Formula of the Mode for Equal Interval Series.....202
6.3.1.2.1.3. Establishing the Mode for Series with Unequal Intervals203
6.3.1.3. Properties of the Mode.....204
6.3.2. The Median.....206
6.3.2.1. Graphic Estimation of the Median for an Interval Series.....212
6.3.2.2. Properties of the Median.....213
6.3.2.3. Quantiles.....214
6.3.3. The Medial.....217

Scanned with OKEN Scanner

7. INDICATORS OF VARIATION AND CONCENTRATION.....223

7.1. The Simple Indicators of Variation.....224
7.2. The Synthetic Indicators of Variation.....226
7.2.1. Properties of Variance and its Simplified Calculus.....236
7.2.2. The Factorial Indicators of Variance.....253
7.3. Indicators of Variance for an Alternative Characteristic.....256
7.4. The Concentration.....257
7.4.1. Properties of the Concentration Curve.....264
7.4.2. The Concentration Index.....266

8. SOME IMPORTANT THEORETICAL DISTRIBUTIONS.....277

8.1. The Binomial Distribution.....277
8.2. The Poisson Distribution.....283
8.3. The Normal Distribution.....285
8.3.1. Properties of the Normal Distribution.....287

9. CHARACTERISTICS OF DISTRIBUTIONS' FORM.....295

9.1. Skewness.....295
9.2. Kurtosis.....305

IV. THEORY AND PRACTICE OF STATISTICAL SAMPLINGS.....307

10. STATISTICAL SAMPLINGS.....309

10.1. Basic Notions, Importance and Types of Samplings.....309
10.1.1. Basic Notions.....309
10.1.2. Representativeness of the Sample.....312
10.1.3. Advantages of Samplings.....314

Scanned with OKEN Scanner

10.1.4. Types of Samplings.....315
10.2. Big Volume Random Samplings.....318
10.2.1. The Simple Random Sampling.....318
10.2.1.1. Sampling with Replacement.....318
10.2.1.2. The Sampling without Replacement.....335
10.2.1.3. The Systematic Sampling.....337
10.2.1.4. Random Sample Size.....338
10.2.2. Cluster Sample.....340
10.2.3. Stratified Sampling.....344
10.2.3.1. Simple Stratified Sampling.....347
10.2.3.2. Proportional Stratified Sampling.....348
10.2.3.3. Optimal Stratified Sampling.....349

11. CHECKING THE STATISTICAL HYPOTHESES.....353

11.1. Notion of Estimation.....353
11.2. Testing Hypotheses.....355
11.2.1. Comparing the Population Mean to a Determined Value 357
11.2.2. Comparing Two Sampling Means.....361
11.2.3. Testing the Standard Error of Median, Quartiles and Deciles 364
11.2.4. Standard Error of the Standard Deviation and of the Variation Coefficient.....366

V. ANALYSIS OF THE CORRELATIONS AMONG PHENOMENA.....371

12. ANALYSIS OF VARIANCE.....373

12.1. Unifactorial ANOVA.....374
12.1.1. Unifactorial ANOVA with Samples of Volume n.....374
12.1.2. Unifactorial ANOVA Based on Samples of different Volume.....382
12.2. Bifactorial ANOVA.....386
12.2.1. Bifactorial ANOVA with One Observation in Each Box.....386
12.2.2. Bifactorial ANOVA with Samples of the Same Volume.....388

Scanned with OKEN Scanner

13. CORRELATIONS AMONG PHENOMENA.....395

13.1. Preliminary Notions.....395
13.2. Elementary Notions for Studying the Relationships among Phenomena.....398
13.2.1. Interdependent Statistical Series' Method.....398
13.2.2. Grouping Method.....399
13.2.3. Graphic Method.....400

14. SIMPLE LINEAR REGRESSION AND CORRELATION.....403

14.1. Simple Linear Regression.....403
14.1.1. Solving Methods of Normal Linear Ecuations Systems.....408
14.2. Duality in Regression.....429
14.3. Intensity of Linear Relationships.....433
14.3.1. Correlation Coefficient.....433
14.3.2. Correlation Rapport.....445
14.3.3. Relationship between Regression and Correlation Coefficient.....457

15. SIMPLE CURVILINIAR REGRESSION AND CORRELATION.....461

15.1. Models of Simple Curviliniar Regression.....461
15.1.1. Parabolic Models.....462
15.1.2. Hyperbolic Models.....476
15.1.3. Power Function Model.....491
15.1.4. Exponential Curve.....496
15.1.5. Logarithm Curve.....504
15.1.6. Deciding the Type of Regression Curve to Fit.....509

Scanned with OKEN Scanner

16. REGRESSION AND MULTIPLE LINEAR CORRELATION.....513

16.1. Basic Notion of Multiple Relationship.....513
16.2. Multiple Linear Regression and Correlation.....514
16.2.1. Multiple Linear Regression.....514
16.2.2. Intensity of Multiple Linear Correlation.....518
16.3. Partial Correlation.....522
16.4. Multiple and Partial Determination.....524

17. MULTIPLE CURVILINIAR REGRESSION AND CORRELATION.....533

17.1. Multiple Curviliniar Regression Equation.....533
17.1.1. Lineation of Multiple Curviliniar Relationship.....536
17.2. Multiple Curviliniar Correlation.....538
17.2.1. Multiple and Partial Coefficient of Determination.....540

18. THE NONPARAMETRIC CORRELATION.....547

18.1. Basic Notion of Nonparametric Correlation.....547
18.2. Spearman's Rank Correlation Coefficient.....548
18.3. Kendall's Rank Correlation Coefficient.....553
18.4. Fechner's Concordance Coefficient.....557

19. THE ASSOCIATION BETWEEN QUALITATIVE CHARACTERISTICS

19.1. Notion of Association.....561
19.2. Existence of the Association.....562
19.3. Intensity of the Association.....570

20. ELASTICITY.....575

20.1.Basic Notion and Types of Elasticity.....575
20.2. Simple Elasticity.....576
20.3. Multiple Elasticity.....582

Scanned with OKEN Scanner

21. THE PRODUCTION FUNCTIONS.....583

21.1. Basic Notions and Types of Production Functions.....583
21.2. One Factor Power Function.....585
21.3. Two Factors Power Functions.....587
21.3.1.The Graphic Presentation of the Two Factors Power Functions.....589
21.3.2. Characteristics of the Production Functions.....591
21.3.3. Average and Marginal Products.....593
21.3.4. Estimation of the Parameters of a Two Variables Power Function.....596
21.4. The Three Factors Power Function.....598

VI. ANALYSIS OF THE PHENOMENA VARIATION IN TIME.....601

22. THE TIME SERIES' INDICATORS.....603

22.1. Absolute Time Series Indicators.....604
22.2. Phenomena Rate of Growth.....611
22.3. Phenomena Rate of Increase.....612
22.3.1. Average Rate of Growth.....614
22.3.1.1. Geometric Mean Method.....615
22.3.1.2. Parabolic Mean Method.....617
22.3.1.3. Exponential Trend Method.....618
22.3.2. Calculus of the Average Rate of Growth for Complex Phenomena.....620
22.4. Forecasting Phenomena on Basis of the Average Growth Indicators.....625
22.4.1. Forecasting Methods Based on the Average Increase.....625
22.4.1.1. Forecasting Absolute Level of a Phenomenon after t Years.....626
22.4.1.2. Forecasting Time Horizon Necessary for a Phenomenon to Attain a Pre-established Level.....626
22.4.1.3. Establishing the Time Period Necessary for Equal Levels of Development of Two Regions A and B.....627

Scanned with OKEN Scanner

22.4.1.4. Equalizing the Development Level of Two Regions for a Pre-established Period of Time.....628
22.4.2. Forecasting Methods Based on the Average Rate of Increase628
22.4.2.1. Calculus of a Phenomenon Level after t Years.....629
22.4.2.2. Establishing the Prognosis Horizon in which a Certain Penomenon Will Attain a Pre-established Level.....629
22.4.2.3. Establishing the Period of Time Necessary for Equalizing the Development Level of Two Regions A and B.....630
22.4.2.4. Equalizing the Development Level of Two Regions after a Pre-established Period of Time.....630

23. LINES AND TIME CURVE FITTING.....633

23.1. The Main Components of Time Series.....633
23.2. Adjustment Methods of Time Series.....637
23.2.1. The Freehand Method.....638
23.2.2. The Method of Moving Average.....639
23.2.3. Method of Semiaverage.....642
23.2.4. The Absolute Average Increase Method.....644
23.2.5. The Average Rate of Growth Method.....647
23.2.6. The Least Squares Method.....649
23.2.6.1. The Linear Trend.....650
23.2.6.2. The Second-Degree Parabola Trend.....655
23.2.6.3. The Third-Degree Parabola Trend.....660
23.2.6.4. The Exponential Trend.....664
23.2.7. Adjustment on a Modified Exponential Curve.....667
23.2.8. The Gompertz Curve.....677
23.2.9. The Logistic Curve.....683
23.2.10. The Second-Degree Exponential Curve.....691
23.3. Interpolation of Time Series.....693
23.3.1. The Lagrange Interpolation Method.....694
23.3.2. The Newton Interpolation Method.....695

24. STUDY OF THE SEASONAL, CYCLICAL AND IRREGULAR MOVEMENTS.....697

24.1. The Notion of Seasonal Movement.....697

Scanned with OKEN Scanner

24.2. Estimation of Seasonal Movement.....702
24.2.1. Analysis of the Seasonality in the Multiplicative Model.....702
24.2.2. Analysis of the Seasonality in the Aditif Model.....711
24.3. Cyclical Movements.....715
24.4. Irregular Movements.....717

VII. INDEX NUMBERS.....719

25. INDEX NUMBERS.....721

25.1. The Notion and Importance of Index Numbers.....721
25.2. Index Classification.....724
25.3. Time Index Numbers.....725
25.3.1. Simple Index Numbers.....726
25.3.2. Sinthetic Index Numbers.....730
25.3.2.1. Aggregative Index Numbers.....730
25.3.2.2. Average Arithmetic Index Numbers.....746
25.3.2.3. Average Harmonic Index Numbers.....747
25.3.2.4. Average Geometric Index Numbers.....748
25.3.2.5. Index Numbers Calculated as a Ratio between Two Arithmetic Weighted Means.....750
25.4. Analysis of the Factors' Influence Using Index Numbers.....757
25.4.1. Using the Aggregative Index Numbers.....757
25.4.2. Using the Average Values' Index Numbers.....762
25.5. Schemes of Index Numbers Decomposition.....767
25.5.1. The Additive Scheme.....767
25.5.2. The Multiplicative Scheme.....769
25.5.2.1. The Method of Separate Influences.....769
25.5.2.2. The Chain Substitution Method.....782
25.6. Territorial Index Numbers.....788
25.7. Plan Index Numbers.....793

APPENDIX.....795
BIBLIOGRAPHY.....797

Scanned with OKEN Scanner

I. PROBLEME ALE STATISTICII CA STIINTĂ

Scanned with OKEN Scanner

Scanned with OKEN Scanner

1. INTRODUCERE

2. NOŢIUNI FUNDAMENTALE ALE STATISTICII

1. INTRODUCERE

1.1. Revoluţia ştiinţifică contemporană şi statistica

Revoluţia ştiinţifică şi tehnică contemporană se caracterizează printr-o *dezvoltare exponenţială* a ştiinţei atât în profunzime, cât şi ca sferă de cuprindere a fenomenelor, ceea ce face ca volumul informaţiilor ştiinţifice să se *dubleze* într-un interval de timp de aproximativ un deceniu. Concomitent cu dezvoltarea şi aprofundarea ei, ştiinţa este supusă şi unui amplu *proces de diversificare*. După estimaţiile unor reputaţi oameni de ştiinţă ar exista în prezent peste 1500 de discipline ştiinţifice. În aceste condiţii, întrepătrunderea dintre ştiinţe este inevitabilă, iar colaborarea dintre ele este o necesitate obiectivă pentru progresul fiecărei ştiinţe. De aceea, cercetarea ştiinţifică în epoca noastră are un pregnant *caracter interdisciplinar*, caracterizat de faptul că la elaborarea uneia şi aceleiaşi teme de cercetare ştiinţifică participă oameni de ştiinţă de specialităţi diferite. *Pierre Anger* [6] arăta că 90% din savanţii şi cercetătorii ştiinţifici care au existat de la începutul dezvoltării ştiinţei sunt acum în viaţă. De la apariţia primei reviste în anul 1665 şi până în prezent, numărul de reviste ştiinţifice şi tehnice a crescut în progresie geometrică.

O caracteristică esenţială a ştiinţei contemporane, care s-a impus în ultimele decenii ca o realitate de maximă importanţă, o constituie *corelaţia* tot mai intensă dintre *dezvoltarea ştiinţei* şi *creşterea economică*. Cuceririle ştiinţei şi tehnicii au determinat mutaţii importante în toate sectoarele de activitate socială şi cu deosebire în sfera producţiei materiale şi a serviciilor. Nivelul învăţământului şi al cercetării ştiinţifice condiţionează astăzi într-o măsură tot mai mare gradul de eficienţă şi de competitivitate al economiei. *Cunoaşterea este putere - knowledge is power* - spunea filozoful englez *Francis Bacon* (1561-1626) acum patru secole.

Potenţialul ştiinţific şi tehnic al unei naţiuni constituie o condiţie importantă a însăşi independenţei ei. Referindu-se la această problemă fundamentală a lumii contemporane, *René Maheu* [120, p.184] arăta că *dobândirea independenţei economice de către naţiuni nu este nicidecum posibilă, dacă potenţialul ei ştiinţific şi tehnic rămâne sub un prag minim*. Fizicianul german *Werner Heisenberg* (1901-1976), unul dintre întemeietorii mecanicii cuantice, membru de onoare al *Academiei Române* şi laureat al *Premiului Nobel* (1932),

Scanned with OKEN Scanner

arăta: "*Întreaga noastră viaţă depinde, fie că suntem de acord fie că nu, de ştiinţă, şi anume într-o măsură pe care mai înainte cu greu ne-am fi putut-o imagina. Cercetarea ştiinţifică nu mai este ornamentul cultural al vieţii - deşi ar putea fi şi aceasta - ci reprezintă sămânţa din care mai târziu vor putea creşte bunăstarea economică, buna organizare a statului, sănătatea politică şi multe altele*"[85, p.211].

Reducerea decalajului ştiinţific dintre naţiuni este cu atât mai importantă, cu cât astăzi se apreciază că el este generatorul tuturor celorlalte decalaje care se manifestă în lumea contemporană.

O altă tendinţă în dezvoltarea ştiinţei şi tehnicii contemporane, cu implicaţii importante în activitatea economică şi socială, o constituie *reducerea* din ce în ce mai pronunţată a *intervalului de timp dintre momentul apariţiei unei descoperiri ştiinţifice şi momentul materializării ei* într-o realizare efectivă. Astfel, de la construirea primei maşini rudimentare, care producea mişcare de rotaţie cu ajutorul aburului, realizată de către matematicianul grec *Heron* din *Alexandria* în sec.I î.e.n. şi până la maşina cu aburi perfecţionată a inventatorului scoţian *James Watt* (1736-1819), brevetată în 1769, s-au scurs aproape două milenii; de la descoperirea principiului fotografiei (1727) până la realizarea ei practică (1839) s-au scurs 112 ani; 56 de ani pentru telefon (1820-1876), inventat de cercetătorul american *Alexander Graham Bell* (1847-1922); 35 de ani pentru radio (1867-1902); 15 ani pentru radar (1925-1940); 12 ani pentru televiziune (1922-1934); 10 ani pentru reactorul nuclear (1932-1942); 6 ani pentru bomba atomică (1939-1945) etc.

Aceste trăsături ale ştiinţei contemporane ridică un grad sporit de dificultate în definirea fiecărei ştiinţe în parte, în delimitarea riguroasă a obiectului lor de studiu, precum şi în precizarea metodelor proprii de cercetare. Unul şi acelaşi fenomen sau ansamblu de fenomene este cercetat de către diferite ştiinţe din puncte de vedere diferite. Însăşi reprezentarea clasică a ştiinţei particulare care, potrivit logicii formale, reprezintă o disciplină ştiinţifică cu un obiect, cu o metodă şi legi proprii, este pusă în discuţie tot mai mult. Una şi aceeaşi ştiinţă poate deveni, într-un anumit context, metodă pentru alte ştiinţe, fiind folosită - alături de alte metode - la realizarea scopurilor ei cognitive. *Caracterul dual* al acestor ştiinţe - de a fi *ştiinţă şi metodă în acelaşi timp* - a fost pus în evidenţă de mai mult timp. Acest caracter îl poartă filozofia, logica, matematica, cibernetica şi statistica. Statistica, fiind o ştiinţă de sine stătătoare, este utilizată tot mai mult în epoca modernă ca o metodă de cercetare de către aproape toate celelalte ştiinţe particulare.

Dezvoltarea rapidă şi diversificată a ştiinţei în epoca contemporană, concretizată în creşterea gradului ei de participare la realizarea procesului de cunoaştere a naturii, societăţii şi gândirii umane, a determinat apariţia şi

Scanned with OKEN Scanner

cristalizarea unor noi ştiinţe, care studiază fie diferite forme de mişcare ale fenomenelor colective, fie unele laturi comune ale acestora, între care se situează şi statistica.

Existenţa obiectivă a unor fenomene de tip colectiv (colectivităţi) ale căror proprietăţi nu pot fi cunoscute decât la nivelul întregului, a impus *statistica* ca o disciplină ştiinţifică indispensabilă pentru *studierea colectivităţilor*.

În condiţiile actuale, când dezvoltarea ştiinţelor sociale cât şi dezvoltarea matematicii au atins un nivel relativ ridicat, procesul matematizării este un proces obiectiv, care contribuie la dezvoltarea şi aprofundarea cunoaşterii ştiinţifice în toate domeniile.

De aceea, folosirea matematicii în ştiinţele sociale şi economice nu trebuie înţeleasă ca un "import" de metode, uneori neadecvate domeniului respectiv, ci trebuie să rezulte din aprofundarea cunoaşterii acestuia, din perfecţionarea mijloacelor şi metodelor ştiinţifice utilizate şi aceasta se realizează prin intermediul statisticii.

1.2. Din istoria statisticii

Folosirea statisticii ca instrument de cunoaştere a vieţii sociale nu este de dată recentă. Ca *formă de evidenţă* a populaţiei şi bunurilor materiale, statistica a fost utilizată din cele mai vechi timpuri. În dezvoltarea ei s-a format ca o disciplină ştiinţifică autonomă într-un îndelungat proces istoric, conturându-şi treptat aria de investigaţii şi perfecţionându-şi în mod continuu noţiunile, categoriile, metodele, procedeele şi tehnicile de cercetare. Ca activitate practică istoriceşte atestată, statistica datează de aproximativ *şase milenii*, iar ca teorie a statisticii - ca ştiinţă - de peste *trei secole*.

Aşa după cum istoria unei ţări nu poate fi scrisă decât bazându-se pe istoria economiei ţării respective, tot aşa istoria economiei nu poate fi scrisă fără a se sprijini pe *istoria statisticii*, singura în măsură să furnizeze şi să analizeze serii de date autentice şi comparabile în timp şi spaţiu, să contribuie, alături de alte ştiinţe, la descoperirea şi descrierea în limbaj propriu a regularităţilor, legităţilor şi legilor economice de-a lungul etapelor istorice succesive pe care le-a parcurs umanitatea. Fără a ne asuma un risc prea mare, putem afirma că statistica unei ţări reprezintă istoria ei în date sau, cum afirma statisticianul german *A.L. Schlozer* (1735-1809) în lucrarea sa *Theorie der statistik* (1804), referindu-se la interacţiunea dintre statistică şi istorie: *"Istoria este statistică în mişcare, iar statistica este istoria în repaus"*.

Unele înregistrări statistice (inscripţii în piatră şi lut ars, papirusuri etc.) care s-au păstrat din antichitate surprind şi astăzi prin ingeniozitatea lor şi este posibil ca noi cercetări statistice şi istorice să dezvăluie faptul că statistica de atunci

Scanned with OKEN Scanner

reprezenta mai mult decât înregistrări discontinue de fapte şi evenimente. Încă din acest stadiu, statistica a servit nevoilor practice ale cunoaşterii şi conducerii vieţii social-economice. Depăşirea stadiului de *statistică practică* sau *materială* s-a produs, ca de altfel şi în alte ştiinţe, aproape în exclusivitate, în momentul în care, alături de interesele practice ale momentului, au apărut în mod necesar şi interese de ordin ştiinţific de cunoaştere critică, de aprofundare, generalizare şi sistematizare a cunoaşterii practice, empirice.

Încă din *antichitate* statistica a fost utilizată pentru descrierea statelor din punct de vedere geografic, economic şi politic. Rădăcinile unei atari concepţii despre statistică se întâlnesc în opera celui mai mare geniu al antichităţii, *Aristotel* (384-321 î.e.n.), cu deosebire în lucrarea sa *Politica*.

În *Evul Mediu* dezvoltarea cunoaşterii statistice a vieţii sociale, economice şi politice ia un avânt deosebit, mai ales ca urmare a creşterii rapide a schimburilor economice dintre state, care necesită informaţii cât mai complete despre condiţiile partenerilor comerciali efectivi sau potenţiali. Printre cele mai reprezentative lucrări de descriere a statelor din epoca Renaşterii menţionăm pe cele elaborate de *Francesco Sansovino* (1521-1583) *Del governo e amministrazione di diversi regni e republiche* şi *Giovani Botero* (1533-1617) *Le relazioni universali*, care au meritul de a fi luat în considerare la descrierea statelor: bogăţiile, populaţia, comerţul, industria şi finanţele lor. Opera lor s-a bucurat de cunoaştere şi continuare şi în alte ţări, printre care şi în *Principatele Române*.

Descrierea statelor, cu caracter empiric până atunci, a devenit disciplină de învăţământ academic guvernată de anumite norme teoretice şi practice, care s-au elaborat şi dezvoltat la început mai ales în universităţile germane, unde s-a format o adevărată şcoală, cunoscută sub denumirea de *Şcoala descriptivă germană*, având ca fondator pe profesorul *Hermann Conring* (1606-1681) de la *Universitatea din Helmstadt*. Prelegerile sale de statistică au fost publicate de către un discipol al lui în anul 1671 sub titlul *Notitia rerum publicarum*. Specialiştii apreciază că cel dintâi *manual de statistică*, considerată drept ştiinţă a statului, aparţine profesorului *Veit Ludwing von Seckendorff* (1626-1692) de la *Universitatea din Hale* şi intitulat *Teutscher Furstenstaat* (1656). Noua disciplină de învăţământ purta denumirea de *Staatskunde* şi ulterior cea de *Statistik*, care se foloseşte până în prezent.

În acelaşi timp cu *Şcoala germană a statisticii descriptive* a luat fiinţă şi s-a dezvoltat în *Anglia* o altă remarcabilă şcoală de statistică numită *Aritmetica politică* după titlul lucrării fundamentale a acestei şcoli *Political Arithmetick, or a Discourse concerning the Extent and Value of Lands, People, Buildings*, elaborată de *William Petty* (1623-1687) între anii 1671-1676 şi publicată la *Londra* în anul 1691. Lucrarea constituie un studiu plin de inovaţii al fenomenelor social-economice prin exprimarea lor cantitativă cu ajutorul *numerelor, greutăţilor şi*

Scanned with OKEN Scanner

măsurilor la nivelul diferitelor state în general, şi al *Angliei* şi *Franţei* în special. Alături de *W.Petty*, alţi reprezentanţi de frunte ai şcolii *Aritmetice Politice* în secolele XVII şi XVIII au fost, în *Anglia, John Graunt* (1620-1674), *Edmund Halley* (1656-1742) si *Gregory King* (1648-1712). În concepţia acestora, fenomenele sociale în general, cele economice şi cele demografice în special, sunt fenomene măsurabile şi pot fi supuse unor prelucrări matematice pe baza cărora se pot desprinde prin analiză regularităţile care le guvernează. Şcoala engleză punea un accent deosebit pe studiul legăturilor cauzale dintre fenomene, pe descoperirea şi descrierea lor numerică. Prin concepţia sa, *Şcoala Aritmeticii Politice* a reprezentat un nou şi important pas înainte în dezvoltarea statisticii ca ştiinţă.

În *Ţările Române*, sub influenţa *Şcolii descriptive germane*, prima şi cea mai reprezentativă lucrare a genului este monografia cărturarului şi domnitorului *Moldovei, Dimitrie Cantemir* - membru al *Academiei de ştiinţe din Berlin* (1714), intitulată *Descriptio antiqui et hodierni status Moldaviae*, redactată în limba latină, în anul 1716, tipărită în limba germană între anii 1769-1770, tradusă apoi în româneşte şi publicată în anul 1825 la tipografia *Mânăstirii Neamţ*.

Această monografie reprezintă o contribuţie românească importantă la dezvoltarea statisticii, pe plan geografic, politic, economic, social şi cultural, care s-a impus în mod deosebit în faţa contemporanilor şi a urmaşilor şi i-a atras autorului faima de savant român de renume european în domeniul statisticii descriptive.

O altă contribuţie românească remarcabilă la dezvoltarea statisticii descriptive o reprezintă activitatea în domeniul acestei ştiinţe a primului economist şi a primului statistician român, om politic în acelaşi timp, *Nicolae Şuţu* (1798-1871), membru al *Royal Statistical Society* din *Londra*. Punctul culminant al activităţii sale de cercetător îl constituie monografia *Notions statistiques sur la Moldavie*, tipărită în limba franceză la *Iaşi* în anul 1849 la editura *A. Henning*, şi după trei ani şi în limba română cu unele revizuiri şi adăugiri. Aceasta a fost urmată în anul 1867 de o altă lucrare intitulată *Quelques observations sur la statistique de la Roumanie* având aceeaşi structură şi caracter ca şi prima sa lucrare. În aceste monografii ca şi în unele manuscrise nepublicate, *Nicolae Şuţu* şi-a asumat sarcina deosebit de complexă şi dificilă de a culege, sistematiza şi interpreta ştiinţific informaţiile statistice veridice provenind din surse diferite în scopul obţinerii unei descrieri statistice adecvate a situaţiei economice din *Moldova* şi *România*. În aceste lucrări se află şi importante consideraţii teoretice ale lui *Şuţu* cu privire la caracterul ştiinţific al statisticii şi la rolul deosebit al acesteia în activitatea de conducere a statelor, ca unul dintre cele mai puternice instrumente de cunoaştere a realităţii social-economice.

În concepţia sa, formată sub influenţa şcolilor franceză şi engleză de specialitate, *statistica este o ştiinţă socială descriptivă, care studiază starea fizică,*

Scanned with OKEN Scanner

morală, intelectuală şi politică a societăţii, îmbrăţişând toate domeniile vieţii sociale, ştiinţa care încearcă să descrie "*faptele variabile*" care sunt determinate de *cauze variabile*.

Alături de *Nicolae Şuţu*, în *Moldova* secolului al XIX-lea s-a afirmat şi agronomul, economistul şi statisticianul *Ion Ionescu de la Brad* (1818-1891), personalitate complexă, care, prin erudiţia şi pluralitatea preocupărilor sale, se înrudeşte pe plan spiritual cu marile figuri ale *Renaşterii* şi care şi-a înscris definitiv numele în galeria oamenilor de ştiinţă români, ca unul dintre întemeietorii statisticii moderne româneşti căreia, prin activitatea sa teoretică şi organizatorică, i-a fixat primele jaloane.

Ion Ionescu de la Brad, învăţat român cu o temeinică pregătire teoretică de economie şi de statistică la *Conservatorul Naţional de Arte şi Meserii* din *Paris*, având ca profesori pe doi dintre cei mai mari statisticieni de la mijlocul secolului trecut (francezul *Alexandre Moreau de Jonnes* (1778-1870) şi belgianul *Adolphe-Lambert Jacques Quételet* (1796-1874)), socotit de unii specialişti ca fiind întemeietorul statisticii sociale moderne, şi-a expus pentru prima dată concepţiile sale în domeniul statisticii în lucrarea *Povăţuiri pentru facerea catagrafiei Moldovei precedate de oarecari elemente de statistică*, publicată în anul 1859. Prima parte a acestei lucrări, *Elemente de statistică*, constituie *primul manual* de statistică elaborat de un autor român, în care savantul expune cu competenţă obiectul şi sarcinile statisticii, metodele, activitatea practică de statistică, organizarea statisticii oficiale, mijloacele de culegere a informaţiilor statistice şi utilitatea statisticii în cunoaşterea realităţilor social-economice ale unui stat.

Concepţia sa despre statistică s-a format sub influenţa şcolii franceze de specialitate. *Ion Ionescu de la Brad* considera *statistica ca fiind o ştiinţă de sine stătătoare care studiază numai faptele sociale măsurabile* şi prin aceasta se detaşează net de concepţia vulgară care prezenta statistica ca o ştiinţă despre treburile statului. Surprind, prin noutatea lor în cadrul epocii, principiile pe care le preconizează în legătură cu recensământul populaţiei (simultaneitatea, periodicitatea, publicarea rezultatelor, precum şi necesitatea întocmirii tabelelor de mortalitate pe medii, sexe şi profesii), care îl plasează pe *Ion Ionescu de la Brad* în rândul celor mai luminaţi gânditori ai vremii în domeniul statisticii şi demografiei.

Aceste principii se vor regăsi, cu acelaşi conţinut, în recomandările *Congresului internaţional de statistică* din anul 1872, care a avut loc la *Petersburg*.

Deşi în expunerea problemelor teoretice ale statisticii s-au strecurat şi unele inexactităţi şi contradicţii, "*Povăţuirile ...*" constituie o lucrare de bază de la începuturile statisticii moderne româneşti, în care este sintetizată experienţa internaţională în această ştiinţă, alături de concluzii desprinse din realităţile româneşti, cu idei originale şi de profunzime.

Scanned with OKEN Scanner

Deşi influenţat în mare măsură de concepţia monografică a lui *Nicolae Şuţu, Ion Ionescu de la Brad* a adus importante *contribuţii creatoare şi în domeniul cercetărilor monografice* pentru studierea vieţii social economice. El elaborează *monografiile judeţelor Dorohoi, Putna şi Mehedinţi*, care au ca scop completarea informaţiilor obţinute dintr-o cercetare totală. Scopul monografiilor de judeţe, mărturisit de autor este de "*a constata starea actuală a agriculturii, de a stabili ce este bun şi folositor şi ... ce vor fi fost odată bune însă care au ajuns a fi păgubitoare*".

Contemporan cu *Nicolae Şuţu* şi cu *Ion Ionescu de la Brad*, la mijlocul secolului trecut, în *Muntenia* s-a afirmat, în domeniul statisticii şi economiei, proeminenta personalitate a lui *Dionisie Pop Marţian* (1829-1865), care a organizat şi condus *primul Oficiu Central de statistică* al acestei provincii în anul 1859 şi ulterior, după *Unire*, pe cel al *Principatelor Unite*, fiind şi *iniţiatorul şi conducătorul primei reviste ştiinţifice de specialitate* din *România*, apărută în anul 1860 şi intitulată *Analele statistice*.

Efectuând studii de specialitate la *Universitatea din Viena* sub îndrumarea profesorului *Lorenz von Stein, Dionisie Pop Marţian* considera *statistica ca o ştiinţă de sine stătătoare, care studiază fenomenele din natură şi societate în scopul descoperirii cauzelor mişcării lor, urmărind formularea legilor* care guvernează dezvoltarea acestor fenomene, *folosindu-se în acest scop de măsurarea şi exprimarea cantitativă a diferitelor lor caracteristici.*

Cu ocazia *Congresului Internaţional de Statistică*, care a avut loc la *Berlin* în anul 1863, *Dionisie Pop Marţian* remarca, cu deplin temei, că "*statistica este singurul tărâm sigur pe care un guvern prevăzător trebuie să întemeieze raţiunea reformelor sale*".

1.3. Cunoaşterea statistică a realităţii

Obiectivul fundamental al statisticii ca ştiinţă este generalizarea experienţei acumulate de practică, elaborarea şi examinarea critică a indicatorilor folosiţi şi a metodelor de calcul şi analiză a acestora, precum şi perfecţionarea lor continuă în vederea îmbunătăţirii practicii în acest domeniu.

Principalul rezultat al activităţii ştiinţifice în domeniul statisticii îl constituie *indicatorii statistici*, fundamentaţi din punct de vedere ştiinţific, care reflectă veridic realitatea. Indicatorii statistici sunt noţiuni induse (obţinuţi prin inducţie). Sensul lor teoretic este inductiv, ceea ce înseamnă că ei trebuie folosiţi întotdeauna în cercetarea şi determinarea realităţii concrete, când atât conţinutul, cât şi forma de manifestare cantitativă a fenomenelor şi proceselor cercetate au un caracter specific şi apar în anumite condiţii bine determinate. Deci, indicatorii

Scanned with OKEN Scanner

statistici sunt noţiuni teoretice elaborate ca rezultat al măsurării şi abstractizării, care trebuie să reflecte cât mai precis anumite trăsături ale realităţii obiective.

Potrivit logicii formale, inducţia şi deducţia sunt două laturi ale procesului unitar al cunoaşterii ştiinţifice în general, ca şi a celei statistice, care se completează şi se condiţionează reciproc.

Progresul ştiinţific în statistică, ca de altfel în oricare altă ştiinţă particulară, se obţine prin parcurgerea a trei etape principale :

1) elaborarea unor generalizări teoretice din ce în ce mai precise şi care pot fi supuse verificării lor în realitate;

2) verificarea teoriilor existente şi reformularea lor tot mai adecvată realităţii pe care o descriu;

3) înlăturarea teoriilor vechi, perimate, şi înlocuirea lor cu altele mai concordante cu cantitatea şi calitatea informaţiilor existente.

Cunoaşterea statistică a fenomenelor complexe se realizează pe baza faptelor constatate dintr-o anchetă, dintr-o serie de experienţe sau dintr-o bază de date, urmărindu-se desprinderea elementelor semnificative din jocul întâmplător al datelor statistice, determinarea unei structuri în aleator sau confruntarea unei teorii cu o realitate. În acest scop statisticienii au elaborat metode şi tehnici tot mai adecvate de obţinere a informaţiilor de verificare a exactităţii lor, de sistematizare, prezentare, de prelucrare, de analiză şi prognoză.

Mersul cunoaşterii ştiinţifice în statistică, de la concret la abstract, înseamnă cercetarea anumitor fenomene concrete sau a anumitor colectivităţi de fenomene, pentru a stabili ceea ce au ele comun, esenţial şi stabil. Pe această cale se ajunge la stabilirea noţiunilor, categoriilor, regularităţilor, legităţilor şi legilor în statistică.

Cunoaşterea statistică *inductivă* rezultă din observarea unor fapte individuale şi izolate sau din observarea unor colectivităţi de fapte şi se exprimă în noţiuni şi legi cu caracter statistic. Aceste noţiuni şi legi generale, obţinute prin inducţia statistică, constituie baza cunoaşterii practice, a determinării, înţelegerii şi transformării realităţii materiale.

Calea inversă a cunoaşterii statistice, care porneşte de la noţiuni şi legi generale (obţinute prin inducţie) către realitatea complexă, constituie a doua latură -cea *deductivă* - a procesului unitar de cunoaştere a realităţii. În acest caz diferite fapte concrete, respectiv colectivităţi de fapte, se reduc la un sistem de indicatori, noţiuni, regularităţi şi legi generale induse.

În felul acesta se desfăşoară procesul cunoaşterii ştiinţifice în statistică, se descriu şi se analizează faptele şi evenimentele cu ajutorul noţiunilor şi legilor generale induse şi devine posibilă explicarea ştiinţifică a realităţii, cunoaşterea tendinţei de dezvoltare a fenomenelor şi luarea de decizii. Deci, generalizările

Scanned with OKEN Scanner

inductive se fac în vederea folosirii în viitor a sistemului de noţiuni şi legi induse pentru cercetarea realităţii concrete.

În general, cercetarea lumii înconjurătoare se desfăşoară începând cu *contactarea directă* a realităţii, moment căruia îi datorăm semnalarea senzorială de noi fapte şi evenimente. *Observaţia simplă* constituie cea mai veche metodă de investigare a realităţii. În forma ei obişnuită, ea este *spontană* şi se exercită asupra fenomenelor care apar întâmplător în câmpul receptivităţii omului. În acest proces *poziţia observatorului este pasivă* şi se rezumă la *înregistrarea faptelor*, iar observaţia apare ca un *proces contemplativ*, în care observatorul rămâne în afara desfăşurării evenimentelor. Acest tip de observaţie are un *caracter empiric* şi duce la *acumularea de experienţă* practică într-un anumit domeniu.

Ştiinţa recurge la un tip mai eficient de observaţie, şi anume la *observaţia ştiinţifică* care se deosebeşte de cea simplă prin *caracterul ei metodic*. Ea devine o *operaţie premeditată* şi organizată, o *căutare sistematică*, care implică întotdeauna *conduita activă* a observatorului.

Obiectul observaţiei îl constituie *diferite însuşiri calitative şi cantitative* ale fenomenelor individuale sau colective şi legăturile dintre ele, precum şi relaţiile dintre fenomene şi ambianţa în care acestea se produc. Uneori este vorba de aspectul formal, structural al fenomenelor, perceptibil cu ochiul liber sau cu instrumente auxiliare, care vizează aspectul cantitativ sau calitativ. În cazul unor fenomene colective, de număr mare, devine necesară observaţia lor selectivă, deoarece capacitatea umană de percepţie este limitată şi nu poate îmbrăţişa simultan întregul complex de detalii necesar cunoaşterii ştiinţifice. De aceea, cercetătorul este obligat să trieze dintre acestea pe cele mai semnificative din punct de vedere al ideii care a generat observaţia, pe care să le supună apoi studiului său atent.

În procesul de cunoaştere complexă a realităţii, prin maniera în care se stabileşte legătura dintre om şi natura înconjurătoare, se diferenţiază două tipuri de observaţie:

- *observaţia directă*, bazată pe percepţia senzorială;
- *observaţia indirectă*, sau mediată.

Desigur, în practică *observăm, de regulă, numai ceea ce cunoaştem,* remarca cu deplin temei scriitorul şi savantul german *Johann Wolfgang Goethe* (1749-1832). Calitatea observaţiei este condiţionată de profunzimea ei. O observaţie intensă permite, chiar dacă vizează un domeniu mai restrâns al realităţii, obţinerea unor informaţii mai importante decât o observaţie mai extinsă, dar mai superficială. Observaţia ştiinţifică implică şi posibilitatea demonstrării constatărilor prin anumite informaţii care constituie elemente cu valoare de argumentaţie. La baza observaţiei se găseşte un proces intelectual complex, întrucât ea sintetizează

Scanned with OKEN Scanner

cele două laturi ale procesului de cunoaştere. Observaţiile noi duc la formularea unor concepte noi, iar noile concepte determină noi observaţii asupra realităţii.

Procesul unitar al cunoaşterii ştiinţifice se poate desfăşura în două forme principale:

a. *Forma simplă*, în care generalizările se obţin plecând de la observarea unor fenomene individuale sau serii de fenomene individuale;

b. *Forma dezvoltată*, în care generalizările se obţin pe baza observării unor colectivităţi de fenomene cu însuşirile lor comune de grup. Forma dezvoltată a procesului de cunoaştere ştiinţifică este *cunoaşterea statistică.*

Când se vorbeşte de metoda statistică de cercetare în realitate, se vorbeşte de inducţie şi deducţie în forma lor statistică. Este clar că inducţia şi deducţia se pot desfăşura fie simplu, fie dezvoltat - statistic.

Forma simplă a procesului de cunoaştere ştiinţifică este un caz particular al formei statistice şi anume atunci când numărul evenimentelor cercetate *n* este egal cu 1, iar acţiunea factorilor neesenţiali - exprimată prin gradul de împrăştiere a nivelurilor individuale în jurul nivelului mediu al evenimentului cercetat - este egală cu zero sau când asupra fenomenului respectiv nu acţionează decât factori esenţiali.

Forma dezvoltată sau statistică a cunoaşterii se referă la cercetarea şi determinarea fenomenelor colective şi a trăsăturilor lor comune. Ea se aplică în cercetarea tendinţelor medii (tipice) şi a cauzelor cu acţiune de masă. În acest caz, numărul fenomenelor cercetate *n* este mai mare decât 1, iar influenţa factorilor neesenţiali asupra unui element oarecare al colectivităţii este mai mare decât zero.

Forma simplă şi cea statistică a procesului de cunoaştere se completează reciproc.

Ceea ce caracterizează în mod specific *obiectul statisticii este faptul că ea studiază fenomenele şi procesele de masă sub formă de mulţimi sau colectivităţi*, care s-au constituit ca rezultat al acţiunii unui mare număr de cauze cu caracter esenţial asociate cu cele întâmplătoare, ale căror influenţe depind de condiţiile concrete de spaţiu şi timp în care se produc. Datorită faptului că fenomenele şi procesele colective depind de un mare număr de factori, forma lor individuală de manifestare poate varia în mare măsură de la o unitate la alta, în funcţie de modul de asociere a influenţei factorilor esenţiali cu cei întâmplători, de intensitatea, stabilitatea, durata şi sensul acestor influenţe. Raportate, însă, la întreaga colectivitate, unităţile individuale, aparent diferite, se prezintă ca asemănătoare între ele din punct de vedere esenţial, datorită faptului că au fost generate de cauze esenţiale comune şi se diferenţiază doar sub aspect cantitativ.

Legităţile care se manifestă în cadrul fenomenelor şi proceselor de masă, în mişcarea lor continuă, nu pot fi scoase în relief, nu pot fi cunoscute şi caracterizate de statistică, decât dacă se iau în studiu toate cazurile individuale sau o parte

Scanned with OKEN Scanner

suficient de mare a lor, care să prezinte la scară redusă trăsăturile principale ale ansamblului, deoarece manifestările individuale sunt extrem de variate.

1.4. Legile statisticii

Se ştie că *legile ştiinţifice* nu exprimă cu certitudine felul cum se comportă natura sau societatea în general, ci ele descriu pur şi simplu modul cum se comportă natura sau societatea între anumite limite şi modul cum probabil se vor comporta din nou în condiţii similare. Statistica, cu ajutorul metodelor proprii, determină care sunt aceste limite, într-un context special de condiţii delimitative şi stabileşte probabilitatea de repartiţie a oricărui ansamblu de evenimente date, pe baza frecvenţei şi regularităţii cu care acestea au apărut în trecut.

Este deosebit de semnificativă ideea că ştiinţele particulare nu sunt cu adevărat dezvoltate decât atunci când ajung să se poată folosi de matematică. Această teză s-a verificat mai de mult în cazul ştiinţelor naturii, iar mai recent şi în cazul ştiinţelor sociale şi economice între care se include şi statistica.

Esenţa statisticii ca ştiinţă, ca teorie, constă în capacitatea ei de a generaliza trăsăturile şi relaţiile dintre fenomenele şi procesele colective din viaţa economică şi socială, care există în mod obiectiv şi în reflectarea lor în noţiuni generale, regularităţi, legităţi şi legi.

Legea statistică este una din formele manifestării legăturilor generale ale fenomenelor din natură şi societate. Spre deosebire de legea dinamică sau de tip dinamic, care se verifică în cazul fenomenelor individuale, luate ca atare (cum ar fi legea căderii corpurilor, legile mecanicii etc.), legea statistică exprimă legătura necesară dintr-un ansamblu de fenomene sau procese, care se găsesc într-o anumită interacţiune, aparent întâmplătoare, şi nu dintre două fenomene oarecare sau două laturi ale unui fenomen. În cazul legilor statistice, fiecare unitate sau fenomen din cadrul ansamblului are o comportare întâmplătoare în raport cu întregul. Legea statistică nu determină mişcarea fiecărui element, ea apare ca rezultantă medie a numeroaselor acţiuni individuale ale ansamblului de fenomene, ca tendinţă predominantă, ca medie a unui număr mare de abateri întâmplătoare.

Între *legile statistice* şi *legile dinamicii* există o legătură dialectică, ele sunt corelate, se întrepătrund, determinând împreună manifestarea fenomenelor, mişcarea lor, tendinţele lor de dezvoltare. Se poate afirma că legile dinamicii sunt cazuri particulare, limită, ale legilor statistice şi anume, atunci când numărul fenomenelor studiate n este egal cu unitatea. Legile sociale acţionează ca legi statistice, realizându-se cu necesitate în ansamblul activităţilor oamenilor.

Astfel, *legea numerelor mari* presupune că într-un număr suficient de mare de cazuri individuale, a căror variaţie este întâmplătoare, influenţa factorilor

Scanned with OKEN Scanner

într-un sens este egal probabilă cu influenţa factorilor de sens contrar şi de aceea se compensează reciproc, în aşa fel încât se obţine o anumită valoare tipică, caracteristică întregii colectivităţi. Este ştiut faptul că fenomenele sociale sunt determinate de o multitudine de cauze esenţiale, prezente în toate cazurile individuale şi care se manifestă cu o anumită intensitate. Dacă asupra fenomenelor respective nu ar acţiona şi factori întâmplători, atunci ar trebui ca, în anumite condiţii determinate, constante, să se obţină aceeaşi manifestare individuală a fenomenelor. În realitate, asupra fenomenelor acţionează şi factori întâmplători, care determină anumite abateri de la ce este normal, esenţial şi tipic, dând naştere la o mare diversitate în lumea fenomenelor sociale colective.

De exemplu, la naştere, proporţia copiilor pe sexe ni se pare întâmplătoare; la nivel naţional însă, şi pe o perioadă îndelungată de timp, se manifestă o lege statistică şi anume la fiecare 100 de fete, se nasc 105-106 băieţi, lege descoperită de englezul *John Graunt* în anul 1662, publicată în lucrarea *Natural and Political Observations upon the Bills of Mortality* şi observată ulterior la aproape toate popoarele lumii.

1.5. Metoda particulară a statisticii

Ştiinţa realizează o unitate indestructibilă a celor două componente esenţiale ale ei: *teoria* şi *metoda*[1] ştiinţifică.

Teoria unei ştiinţe constă din fondul de cunoştinţe acumulate, transpuse într-o formă raţionalizată de concepte şi legi, iar metoda este modul de cercetare, de cunoaştere a realităţii obiective, sau mai simplu, modul de a opera cu fondul de cunoştinţe existent pentru a putea dezvolta în continuare cunoaşterea. Aceste două componente ale oricărei ştiinţe particulare sunt inseparabile.

Teoria ştiinţifică ar stagna la un anumit nivel, dacă progresul în cunoaştere nu ar mai fi posibil. Pe de altă parte, metoda izolată de contextul informaţional al ştiinţei ar deveni improprie pentru noi descoperiri şi şi-ar pierde utilitatea. Tocmai metoda, ca unică modalitate de stabilire a adevărului ştiinţific, este elementul care conferă ştiinţei mobilitate şi perfectibilitate permanentă.

Metoda ştiinţifică, arăta statisticianul şi filozoful englez *Karl Pearson* (1857-1936), constă în clasificarea atentă şi migăloasă a faptelor, în compararea relaţiilor dintre ele în mod succesiv şi, în final, în descoperirea pe baza imaginaţiei creatoare a unei relaţii sau a unei formule, care în câteva cuvinte (simboluri)

[1] Din grecescul "*methodos*" care înseamnă *drum, cale, mod de expunere.*

Scanned with OKEN Scanner

rezumă un mare număr de fapte. În ştiinţă o astfel de formulă poartă denumirea de lege [2].

În activitatea de cercetare, fiecare disciplină ştiinţifică foloseşte un ansamblu de metode şi tehnici. Unele îi sunt specifice, iar altele sunt preluate de la disciplinele înrudite. Ansamblul metodelor folosite într-o anumită ştiinţă reprezintă *metodologia particulară* a ştiinţei respective. Acestea se situează la un nivel subordonat metodologiei generale a oricărei ştiinţe. Însuşirea şi perfecţionarea metodologiei ştiinţifice este de o importanţă deosebită pentru fiecare ştiinţă particulară.

Ceea ce caracterizează sub aspect metodologic statistica este exprimarea fenomenelor şi proceselor colective sub *formă cantitativă, numerică.*

Conceptul de *număr* a apărut în antichitate în urma unui îndelungat şi anevoios proces de abstractizare a rezultatelor unor observaţii empirice a fiinţelor şi lucrurilor, a faptelor şi evenimentelor, care se iveau în câmpul receptivităţii omului. Introducerea simbolurilor literale, deci a unităţilor imaginare, în locul numerelor a reprezentat o nouă treaptă ascendentă în procesul de abstractizare care a contribuit în mod hotărâtor la constituirea statisticii ca ştiinţă.

Este deosebit de sugestivă, în acest sens, reflecţia poetului şi aviatorului francez *Antoine de Saint-Exupery* (1900-1944), dintr-un poem al său: "*On ne voit bien qu'avec le coeur, l'essentiel est invisible pour les yeux*".

Apariţia conceptului de număr reprezintă în statistică un pas hotărâtor în procesul de trecere de la lumea fenomenelor reale, empirice, perceptibile pe cale senzorială directă, la domeniul structurilor ideale, raţional conceptuabile, cu mare putere de generalizare. Strâns legate de conceptul de *număr*, au apărut ulterior şi conceptele de *numărare, calculare* şi apoi cel de *măsurare*. Conceptul de număr l-a generat pe cel de numărare, iar acesta a determinat apariţia şirurilor de numere în matematică şi mai târziu a conceptelor de grupuri, colectivităţi, variabile, serii şi distribuţii în statistică.

În antichitate, conceptul de bază al ştiinţei, după *Aristotel*, era: "*lucrurile sunt numere*" sau "*cerul întreg este armonie şi număr*" ca rezultat al ordinii şi armoniei obţinute din abstractizare.

Propoziţia *"unu plus unu fac doi"* este adevărată, indiferent de natura fenomenelor care se adună, pe baza puterii de ordonare a conceptului de număr abstract, care apoi a fost înlocuit cu simboluri literale.

Abstractizarea este o cale a gândirii conceptuale care se bazează pe analiza trăsăturilor calitative, pe ierarhizarea şi trierea acestora în scopul separării şi luării în considerare numai a unei proprietăţi determinate a lor, neglijând pe toate

[2] Pearson Karl - *The Grammar of Science*, Adam and Charles Black, London, 1900, p.77.

Scanned with OKEN Scanner

celelalte. Abstractizarea contribuie la descoperirea trăsăturilor comune ale fiinţelor şi lucrurilor, faptelor şi evenimentelor, care se constituie în acte de cunoaştere, având o semnificaţie deosebită. *Conceptualizarea* se întemeiază pe procesul abstractizării şi permite relevarea a ceea ce este echivalent între fenomene, deoarece egalitatea lor deplină, absolută, nu apare în realitate niciodată.

Raportul dintre un fenomen şi structura sa de bază este deosebit de complex şi nu poate fi urmărit cu precizie în fiecare caz individual, particular. Se poate aprecia că până în prezent doar ştiinţele naturii au reuşit într-o măsură mai mare clarificarea relaţiei dintre conceptele cu care se descriu fenomenele naturii şi acelea simplificatoare, care apar în formularea legilor naturii.

Ştiinţa progresează, se dezvoltă continuu, pentru că slujitorii ei îşi pun mereu noi întrebări la care caută răspunsuri tot mai adecvate, deoarece ei consideră lumea un sistem unitar cu o structură şi cu o ordine internă greu de înţeles, elaborând în acest fel concepte tot mai cuprinzătoare.

Se poate afirma că oamenii de ştiinţă dialoghează în mod continuu cu natura înconjurătoare, pentru a-i cunoaşte mai bine esenţa, structura şi ordinea ei internă şi pentru a trăi mai bine în ea.

Caracteristicile cantitative sunt transformate în studiu statistic sub formă cantitativă prin *cuantificare*. Acest fapt creează posibilitatea prelucrării şi analizei cantitative a fenomenelor colective, simplificării şi caracterizării în mod sintetic şi generalizat a particularităţilor lor calitative.

După cum se ştie, lumea fenomenelor naturale este cognoscibilă pe cale experimentală şi raţionalizabilă. În procesul de cunoaştere a fenomenelor sociale şi economice experimentul nu este decât rareori posibil şi cel mai adesea el se înlocuieşte cu observaţia faptică.

Observaţia ştiinţifică faptică - ca modalitate de cunoaştere a realităţii - nu se limitează numai la constatarea şi înregistrarea faptelor şi evenimentelor, ci se completează în mod necesar cu raţionalizarea acestora cu ajutorul metodelor statistice, utilizând instrumentul matematic.

În cadrul observaţiei, instrumentul principal îl constituie dimensionarea, măsurarea fenomenelor şi proceselor supuse cercetării. Conceptele fundamentale de mărime măsurabilă, procedeu de măsurare şi rezultat al măsurării, care poate fi exprimat în numere întregi sau fracţionare, au apărut pentru prima dată în *Geometrie.*

Operaţia de măsurare reprezintă un proces experimental de obţinere a informaţiei sub forma unui raport numeric între valoarea mărimii (fizice) măsurate M şi valoarea unei alte mărimi considerate drept unitate de măsură m:

$$n = \frac{M}{m} \qquad (1.1)$$

Scanned with OKEN Scanner

Din această expresie, numită uneori şi *ecuaţia fundamentală a măsurării*, se poate constata că rezultatul final presupune efectuarea prealabilă a operaţiilor de definire şi descriere a obiectului măsurării, de codificare şi prelucrare a informaţiilor de măsurare.

În calitate de obiect al măsurării nu apar lucrurile însele, faptele şi evenimentele existente în realitate, ci anumite proprietăţi ale acestora şi ceea ce măsurăm direct nu constituie proprietăţi obiectiv existente, independente de procedeul de măsurare, ci reprezintă doar proprietăţi construibile, atribuabile, concordante cu nivelul de dezvoltare a ştiinţei, pe baza cărora se încearcă a se caracteriza ceea ce se află îndărătul caracteristicilor măsurate. Aceasta înseamnă că măsurarea nu este decât o modalitate concretă de definire, descriere şi cunoaştere cantitativă a obiectelor supuse măsurării.

Măsurarea şi, pe această bază, obţinerea indicatorilor cantitativi şi calitativi ai fenomenelor sociale şi economice, descoperirea şi stabilirea legăturilor dintre ei, permite trecerea la o etapă superioară a cunoaşterii şi anume, la modelarea statistică a realităţii.

Practica a arătat că măsurările, indiferent de natura lor şi oricât de corect s-ar efectua, sunt afectate în mod obligatoriu de erori, datorită variaţiei în timp şi spaţiu a obiectului sau fenomenului măsurat, imperfecţiunii organelor noastre de simţ, aparaturii şi metodelor de măsurare, precum şi influenţei condiţiilor exterioare. Evident, prin perfecţionarea aparaturii şi a metodelor de măsurare se pot obţine aproximaţii din ce în ce mai bune ale entităţilor observate, însă mărirea preciziei măsurării este limitată.

Diferenţa dintre rezultatul obţinut prin măsurarea unui anumit obiect sau fenomen şi valoarea lui reală poartă denumirea de *eroare de măsurare*:

$$\Delta_{x_i} = x_i - X \qquad i = \overline{1,n} \tag{1.2}$$

unde: x_i - rezultatul măsurării i;
X - valoarea reală a fenomenului măsurat;
Δ_{x_i} - eroarea de măsurare;
n - numărul de măsurări.

Desigur, erorile de măsurare intervin în orice proces de măsurare, indiferent dacă aceasta se face în natură sau în societate. În general, se consideră că erorile de măsurare sunt mai mari în cazul fenomenelor sociale decât cele corespunzătoare fenomenelor naturii, datorită unei variabilităţi şi a unei diversităţi mai mari ale acestora, ca şi datorită inexistenţei unor mijloace de măsurat adecvate.

Statistica şi-a elaborat metode specifice de estimare a erorilor de măsurare.

Scanned with OKEN Scanner

În orice cercetare statistică concretă şi completă se pot identifica în general patru etape principale:

1. *Obţinerea informaţiilor* statistice vechi şi noi, necesare cu privire la fenomenele cercetate;

2. *Prelucrarea multilaterală* a acestora pentru a determina indicatorii sintetici necesari caracterizării sub diferite aspecte a fenomenelor sau proceselor studiate;

3. *Analiza indicatorilor statistici*, obţinuţi prin prelucrare, pentru stabilirea şi măsurarea legăturilor dintre fenomene, pentru a se ajunge la explicaţia lor ştiinţifică şi la generalizările teoretice sub formă de legi, legităţi şi regularităţi statistice;

4. *Prognoza fenomenelor* efectuată pe baza rezultatelor analizei statistice în vederea folosirii lor în procesul decizional.

1.6. Definiţia statisticii

Termenul *statistică* derivă din latinescul *status* cu sensul pe care l-a avut în latina medievală de *stare politică*.

Accepţiunea actuală a cuvintelor *statistică*, *statistician*, *statistic* datează de peste 100 de ani. Totuşi aceste cuvinte au o vechime mult mai mare şi sensul lor a suferit modificări în timp.

Semnificaţiile actuale ale cuvântului *statistică* sunt următoarele:

a) *Date numerice* cu privire la unul sau mai multe fenomene din natură sau din societate, culese, sistematizate, centralizate şi uneori publicate în anuare statistice, culegeri de date statistice etc.;

b) *Activitate de culegere, prelucrare şi valorificare a informaţiilor statistice*, organizată de stat sau individual de către cercetători;

c) Totalitatea metodelor, procedeelor şi tehnicilor statistice sau *metodologia statistică*;

d) *ştiinţă de sine stătătoare*.

În raport cu alte ştiinţe, în care este folosită, statistica este o metodă de cercetare.

Ţinând seama de obiectul şi metoda ei de cercetare vom defini statistica ca fiind *o ştiinţă metodologică care studiază fenomenele şi procesele colective din natură şi din societate pe baza observării lor ştiinţifice, a culegerii şi prelucrării detaliate a informaţiilor statistice cu scopul de a formula şi explica legile, legităţile şi regularităţile ce le guvernează*.

În literatura de specialitate se întâlnesc numeroase definiţii ale statisticii, unele chiar umoristice, altele confuze, unele prea largi, altele prea limitate. O

Scanned with OKEN Scanner

definiţie unanim acceptată nu există încă. Poate studiul evoluţiei conţinutului acestei noţiuni de-a lungul timpului ar putea duce la unele clarificări.

În condiţiile actuale are loc o lărgire considerabilă a cadrului de aplicare a metodelor statistice moderne în toate domeniile de cercetare ştiinţifică, ca şi în rezolvarea unor probleme practice ale etapei de tranziţie spre economia de piaţă în ţara noastră, acordându-se o atenţie deosebită statisticii şi aplicaţiilor acesteia în viaţa social-economică. Acest fapt este de la sine înţeles dacă ţinem cont că metodele statistice sunt folosite, în special, la elaborarea sistemului informaţional economic şi social, la elaborarea prognozelor sociale, economice şi tehnologice, ca o bază ştiinţific fundamentată, pentru elaborarea programelor curente şi de perspectivă ale dezvoltării economico-sociale. De aici rezultă că sfera celor interesaţi în cunoaşterea noţiunilor fundamentale ale statisticii şi a metodelor, care îşi găsesc aplicaţii eficiente în activitatea practică de producţie, se lărgeşte din ce în ce mai mult cuprinzând, pe lângă statisticieni şi economişti, pe demografi, sociologi, ingineri, tehnicieni etc., care lucrează în cele mai variate domenii ale vieţii sociale şi economice.

La nivel de întreprindere, *Statistica* reprezintă un instrument util pentru îmbunătăţirea administrării activităţii acestora, permiţând exercitarea unui control permanent asupra procesului productiv de bunuri sau de servicii, ca şi în fundamentarea unor măsuri corespunzătoare în cazul unor perturbaţii importante care pot să apară în timp, cum ar fi: variaţii importante ale cererii sau ale ofertei, insuficienţa temporară a unor resurse materiale, de energie etc. Asemenea fapte necesită o bună cunoaştere a fenomenelor care trebuie conduse. Statistica este aceea care permite administratorului să-şi structureze şi sintetizeze informaţiile potenţial disponibile şi să le utilizeze în fundamentarea deciziilor celor mai adecvate.

Volumul informaţiilor disponibile în întreprinderi este important şi se referă la specializarea diferitelor ei compartimente ca şi relaţiile acesteia cu alte întreprinderi care necesită numeroase schimburi reciproce de informaţii utilizate în scopuri de coordonare: note de comandă, note de livrare, facturi, note contabile, fişe ale utilajelor, fişe de salarii etc., ţinute pe suporturi fizice sau informatice. Toate aceste forme de evidenţă sunt întocmite cu scopul precis de a determina drepturile şi obligaţiile legale ale întreprinderii faţă de furnizori, de clienţi şi de proprii salariaţi.

În funcţie de sursa lor, informaţiile necesare administratorului se pot împărţi în două categorii şi anume în *informaţii interne* arhivate temporar, care, după ce au fost utilizate se distrug sau sunt centralizate pentru a servi o durată mai mare de timp şi *informaţii externe* întreprinderii, ţinute de instituţii specializate cum ar fi *Comisia Naţională pentru Statistică*.

Scanned with OKEN Scanner

Unii cercetători din domeniul ştiinţelor sociale folosesc statistica în studiile lor în mod neadecvat, la fel cum foloseşte un om beat stâlpii de iluminat de pe străzi, mai mult drept punct de sprijin decât pentru iluminat.

"*Fără o cunoaştere adecvată a statisticii* - se arată într-un curs de mare circulaţie în S.U.A. - *cercetătorul din domeniul ştiinţelor sociale se găseşte în situaţia obişnuită a unui om lipsit de vedere care caută într-o cameră întunecată o pisică neagră care nu este acolo*".[3]

Putem afirma cu toată certitudinea că, dacă (atât în trecut, cât şi astăzi) este unanim acceptat faptul că legile naturii sunt scrise în limbaj statistic, legile sociale şi cele economice - mai devreme sau mai târziu - vor fi scrise în acelaşi limbaj şi aceasta este o problemă fundamentală care stă în faţa ştiinţelor sociale în general, şi a statisticii în special.

Acolo unde nu este adevăr, nu poate fi nici ştiinţă, celor ce se tem de adevăr le este contraindicat să se ocupe de ştiinţă; a semăna minciuna în informaţia statistică înseamnă a discredita această ştiinţă şi în final a aduce prejudicii societăţii.

Se poate afirma cu deplin temei că în epoca contemporană metoda caracteristică a cercetării ştiinţifice moderne, aproape în toate domeniile, este metoda statistică. Statistica asigură baza sistemului informaţional modern, iar deciziile de ordin politic, social şi economic pe plan naţional, regional şi local, ca şi în întreprinderi se bazează pe informaţia statistică.

---

[3] Traducerea noastră din: Croxton F.E., Cowden D.J., Klein S. - *Applied General Statistics*, Preutice-Hall Inc., Englewood Cliffs, N.J., 3rd edition, 1967, p.1: *Without an adequate understanding of statistics, the investigator in the social sciences may frequently be like the blind man groping in a dark closet for a black cat that isn't there.*

Scanned with OKEN Scanner

2. NOŢIUNI FUNDAMENTALE ALE STATISTICII

2.1. Colectivităţi statistice

Fenomenele şi procesele din natură şi societate se prezintă obişnuit sub forma unor *grupuri, ansambluri sau mulţimi* de elemente individuale de acelaşi fel, care au nevoie pentru apariţie şi dezvoltare de condiţii asemănătoare şi urmează aceleaşi legi. Atunci când o asemenea mulţime de fenomene constituie obiectul cercetării statistice, ea poartă denumirea de *colectivitate statistică, populaţie*[1] sau *fenomen de masă*.

Colectivităţile statistice pot fi definite prin *extensie* sau prin cuprindere, enumerând elementele (unităţile) care le compun ori prin *înţelegere*, pornind de la un criteriu sau de la o proprietate de apartenenţă pe care o posedă toate unităţile lor componente.

Colectivitatea statistică reprezintă deci un ansamblu de fenomene individuale, care au una sau mai multe caracteristici esenţiale comune, bine definite, proprii tuturor unităţilor ei componente. Colectivităţile se caracterizează prin existenţa lor de sine stătătoare, bazată pe asociaţie, pe legături între elementele lor componente. Statistica cercetează *colectivităţi finite* a căror *existenţă este obiectivă şi concretă sau faptică*. Exemple de colectivităţi statistice sunt numeroase: populaţia unei ţări sau a unui judeţ, formată din persoane legate între ele prin calitatea de a fi cetăţean al ţării sau al judeţului; industria compusă din întreprinderi industriale, legate între ele prin caracteristica de a fi producătoare de bunuri care satisfac anumite necesităţi; producţia unei întreprinderi, formată din diferite produse cu caracteristici asemănătoare; locuinţele dintr-un oraş sau comună; numărul căsătoriilor dintr-un anumit teritoriu care s-au încheiat într-o perioadă de timp determinată; numărul naşterilor care au avut loc la o anumită populaţie într-o perioadă de timp determinată etc. Din aceste exemple rezultă

[1] Denumirea de "*populaţie*" îşi are originea în faptul că primele colectivităţi concrete care au fost studiate de către statistică se refereau la populaţia omenească.

Scanned with OKEN Scanner

că în viaţa socială *colectivităţile pot fi formate din fiinţe, din lucruri şi din manifestări ale acestora (fapte şi evenimente).*

Acţiunea aceloraşi legi asupra tuturor elementelor unei colectivităţi determină gradul lor de *omogenitate*. Elementele colectivităţii sunt legate între ele prin una sau mai multe trăsături esenţiale comune, însoţite de altele neesenţiale, care se diferenţiază de la o unitate la alta sau de la un grup de unităţi la altul, care dau colectivităţii un caracter de complexitate.

Însuşirea de bază a colectivităţilor este *omogenitatea*. O colectivitate statistică este omogenă dacă elementele ei componente au trăsături esenţiale asemănătoare. Desigur, omogenitatea nu este sinonimă cu identitatea. Statistica, în scopul de a caracteriza o anumită colectivitate, face abstracţie de diversitatea elementelor ei individuale, luând în consideraţie numai caracteristicile de bază ale acestora.

În cadrul unor colectivităţi se pot individualiza grupuri mai mult sau mai puţin numeroase de unităţi, *subcolectivităţi* (colectivităţi mai restrânse) în funcţie de nivelul de dezvoltare a caracteristicilor principale. Într-o colectivitate dată, subcolectivităţile se caracterizează totdeauna printr-un grad mai mare de omogenitate decât colectivităţile din care provin.

O cercetare statistică nu se poate limita numai la studiul trăsăturilor generale ale colectivităţii ca întreg, ci trebuie să meargă mai departe, în profunzime, şi să dezvăluie anumite tendinţe de grup, tipice sau a *tipurilor calitative din cadrul lor*, contribuind astfel la cunoaşterea lor structurală, veridică, completând cunoaşterea colectivităţii ca întreg şi permiţând stabilirea legăturilor dintre tipuri în cadrul ansamblului. Tipurile calitative ale unei colectivităţi pot apare sub forma unor grupe de unităţi statistice mai mult sau mai puţin numeroase. Uneori tipul se referă la o categorie de unităţi pe cale de a se forma - *noul* - alteori el dezvăluie o categorie pe cale de dispariţie - *vechiul*.

Omogenitatea colectivităţilor poate fi privită sub trei aspecte principale: *din punct de vedere calitativ, de timp şi de spaţiu.*

Omogenitatea calitativă a colectivităţilor este dată de însuşirea elementelor componente de a fi de acelaşi gen din punct de vedere al însuşirii sau însuşirilor esenţiale (calitative) care le definesc. De exemplu, în funcţie de sex, populaţia unei ţări se poate considera că este formată din două subcolectivităţi mai omogene: populaţia feminină şi populaţia masculină; aceeaşi colectivitate după vârstă se poate împărţi în trei subcolectivităţi mai omogene: *tineri, adulţi şi vârstnici*; din punct de vedere al stării

Scanned with OKEN Scanner

civile se pot distinge patru subcolectivităţi: *necăsătoriţi, căsătoriţi, văduvi şi divorţaţi* etc.

Omogenitatea de timp este însuşirea elementelor componente ale colectivităţilor de a aparţine aceluiaşi moment sau aceleiaşi perioade de timp. Condiţiile momentului sau ale perioadei îşi imprimă influenţa asupra gradului de dezvoltare a elementelor componente ale colectivităţii. De exemplu, colectivitatea produselor dintr-o zi ale unei fabrici, populaţia unei ţări obţinută pe baza unui recensământ, numărul locuinţelor înregistrate cu ocazia unui recensământ.

Omogenitatea de spaţiu (teritorială) este însuşirea elementelor componente ale unei colectivităţi de a aparţine unui anumit teritoriu, care imprimă anumite trăsături colectivităţii respective.

În vederea înregistrării statistice a unei colectivităţi, este necesară definirea şi delimitarea ei cât mai riguroasă din punct de vedere calitativ, în timp şi în spaţiu.

În viaţa socială şi economică colectivităţile sunt numeroase şi de aceea, pentru studiul lor sistematic, este necesar să fie clasificate în funcţie de criterii diferite.

După natura lor, colectivităţile se pot împărţi în două categorii: *colectivităţi de stări (de stocuri) şi colectivităţi de mişcări (de fluxuri)* [14,p.44-45].

Colectivităţile de fiinţe sau de lucruri care exprimă starea (stocul) acestora la un anumit moment poartă denumirea de *colectivităţi de stări*. Aceste colectivităţi se caracterizează prin esenţa lor materială, concretă, statică şi prin faptul că volumul lor poate fi determinat prin numărare sau măsurare la un moment dat.

Colectivităţile de fapte sau de evenimente, numite şi *colectivităţi de mişcări (de fluxuri)*, reprezintă manifestări ale colectivităţilor de fiinţe sau de lucruri şi se caracterizează prin esenţa lor nematerială, faptică şi prin aceea că volumul lor nu poate fi determinat prin numărare sau măsurare la un moment dat, ci prin înregistrarea curentă continuă a evenimentelor (faptelor, ideilor, opiniilor) care se produc într-un interval de timp dat şi într-un spaţiu determinat.

Scanned with OKEN Scanner

Această clasificare este importantă în statistică, deoarece în funcţie de natura colectivităţilor, se alege modalitatea concretă de înregistrare şi de prelucrare a informaţiilor cu privire la colectivităţile cercetate.

Din punct de vedere al gradului de cuprindere a unităţilor care se înregistrează, colectivităţile se împart în două categorii: *colectivităţi generale* şi *colectivităţi parţiale*.

Colectivităţile generale (totale) se obţin prin înregistrarea tuturor unităţilor supuse observării statistice totale. Dacă nu se face o altă precizare atunci, ori de câte ori se vorbeşte despre colectivităţi statistice, se subînţeleg colectivităţi generale.

Colectivităţile parţiale (de selecţie) se obţin ca rezultat al înregistrării unei părţi din colectivităţile generale prin cercetări selective, anchete etc.

2.2. Unităţi statistice

Elementele componente ale colectivităţilor care se înregistrează cu însuşirile lor poartă denumirea de *unităţi statistice*. Ele sunt supuse observării statistice şi înregistrării cu datele corespunzătoare însuşirilor lor, care interesează cercetarea. De aceea, unitatea statistică mai poartă şi denumirea de *unitate de observare*, care trebuie clar *definită, identificată* pe teren, *înregistrată* şi apoi *analizată*.

Unităţile statistice sunt fapte elementare care fac obiectul observaţiei indiferent de natura lor, cu condiţia de a răspunde cerinţelor unei definiţii precise. Ele pot fi fiinţe, lucruri, fapte, evenimente, opinii, comportamente, obiceiuri etc.

Din punct de vedere al numărului de caracteristici necesare pentru exprimarea şi descrierea lor, unităţile statistice se pot împărţi în două categorii:

- unităţi statistice care se pot exprima (defini) pe baza unei singure caracteristici (lungime, lăţime, suprafaţă, greutate, volum, intensitate etc.);
- unităţi statistice care se exprimă prin luarea în consideraţie a două sau mai multe caracteristici (de exemplu timpul de muncă se poate exprima prin *om-zile, om-luni, om-ani*; activitatea de transport se poate exprima în *tone-kilometri* parcurşi; presiunea se exprimă în *kg* sau grame pe unitatea de suprafaţă; viteza unui vehicul se exprimă în *kilometri pe unitatea de timp* etc.

După componenţa lor se disting două feluri de unităţi statistice: *simple* şi *complexe*.

Unităţile statistice simple sunt formate dintr-un singur element indivizibil, cum ar fi persoana în cadrul unui recensământ al populaţiei, o naştere, un deces, un produs al unei fabrici, un act de vânzare, de cumpărare, de contractare, o operaţie de prelucrare la un produs etc.

Scanned with OKEN Scanner

Unităţile statistice complexe sunt formate din două sau mai multe unităţi simple, constituite ca rezultat al organizării sociale şi economice a colectivităţii, de exemplu: familia formată din mai multe persoane, echipa şi brigada de lucru, grupa de studenţi, lotul de produse, clădirea formată din mai multe apartamente, biblioteca etc.

Definirea clară şi precisă a unităţilor statistice este o condiţie de bază pentru ca într-o cercetare statistică să se ajungă la înregistrări de *date comparabile*. Definiţia unică pe întreg arealul cuprins în cercetare stă la baza comparaţiilor ce trebuie făcute între aspectele diferite pe care le îmbracă un anumit fenomen în teritorii diferite, în funcţie de condiţiile specifice ale dezvoltării.

Totalitatea unităţilor statistice care fac parte din cadrul unei colectivităţi formează *volumul acesteia. Volumul colectivităţilor* se determină prin identificare, înregistrare, numărare şi totalizare a unităţilor care aparţin prin definiţie colectivităţilor respective.

2.3. Caracteristicile unităţilor statistice

Însuşirea sau trăsătura comună tuturor unităţilor statistice ale unei colectivităţi, ale cărei valori diferă, în general, de la o unitate statistică la alta sau de la un grup de unităţi la altul, poartă denumirea de caracteristică statistică.

Fiecare unitate statistică a unei colectivităţi are în mod obişnuit mai multe *caracteristici*, dintre care unele prezintă interes în cercetările concrete.

Astfel, dacă se studiază colectivitatea personalului unei întreprinderi pot fi avute în vedere caracteristicile: sexul, vârsta, nivelul de calificare, salariul lunar etc. Unele caracteristici se referă la natura internă a unităţilor, iar altele la nivelurile pe care diferite unităţi l-au atins în dezvoltarea lor.

Caracteristicile unităţilor statistice se pot clasifica din mai multe puncte de vedere.

a) După *natura lor*, caracteristicile se pot împărţi în trei mari categorii: *de timp, de spaţiu şi calitative.*

Caracteristicile cantitative reprezintă rezultatul unei măsurări pentru care se pot defini relaţiile de egalitate şi însumare.

Scanned with OKEN Scanner

O caracteristică cantitativă poate fi privită (sau transformată) calitativ în scopul simplificării şi sintetizării realităţii. De exemplu, mulţimea salariilor individuale dintr-o întreprindere se poate împărţi în trei categorii: salarii mici, salarii medii şi salarii mari. Transformarea inversă a unei caracteristici calitative în una cantitativă nu se poate realiza.

Diverse *stări* pe care le pot poseda caracteristicile calitative şi diferite valori numerice pe care le pot lua caracteristicile cantitative se numesc *variante* sau *modalităţi* ale lor.

Măsurarea unei caracteristici pe care o posedă o unitate statistică constă în atribuirea unui *scor*, care arată în ce grad ea posedă o proprietate corespunzătoare caracteristicii studiate prin utilizarea unei *scări de măsură* care prezintă proprietăţi izomorfe cu cele ale caracteristicii studiate.

Variabilele asupra cărora se poate aplica acelaşi ansamblu de operaţiuni de prelucrare statistică sunt *variabile măsurabile* pe acelaşi tip de scară.

Se cunosc şi se utilizează în practică *patru categorii de scări de măsurare*, care se prezintă în ordine descrescătoare a posibilităţilor lor de măsurare astfel:

- *Scări proporţionale;*
- *Scări de intervale;*
- *Scări ordinale;*
- *Scări nominale.*

Informaţiile statistice obţinute pe baza *scării nominale* permit caracterizarea tendinţei centrale cu ajutorul *modului* şi studiul asociaţiei dintre caracteristicile calitative.

Scara ordinală dă posibilitatea calculării, în plus faţă de cea nominală, şi a cuantilelor şi a corelaţiei rangurilor.

În condiţiile utilizării *scării de intervale* se pot calcula în plus faţă de scara ordinală indicatorii tendinţei centrale, indicatorii dispersiei, regresiei şi corelaţiei şi se pot aplica diferite teste parametrice.

Pe baza informaţiilor provenite din utilizarea *scării proporţionale* se poate calcula în plus şi media geometrică şi coeficientul de variaţie.

Scara proporţională este cea mai puternică, iar cea nominală este cea mai puţin restrictivă.

Scara ordinală conţine toate informaţiile pe care le cuprinde scara nominală. Entităţile considerate echivalente pe o scară nominală sunt în mod necesar de acelaşi rang pe o scară ordinală.

Scanned with OKEN Scanner

b) După *importanţa lor* în procesul cunoaşterii statistice, caracteristicile pot fi: *esenţiale*, care definesc colectivitatea şi sunt prezente la toate unităţile componente, şi *neesenţiale* sau cu caracter întâmplător.

c) În funcţie de *numărul de valori* (niveluri sau variante) pe care le pot lua, caracteristicile se împart în două mari grupe: caracteristici *alternative* (sau binare) şi caracteristici *nealternative*.

Caracteristicile alternative sunt acele însuşiri ale unităţilor statistice care nu pot lua decât două valori, de exemplu: sexul (bărbat, femeie), mediul social (rural, urban), calitatea producţiei (bună, rebut).

Caracteristicile nealternative reprezintă trăsături ale unităţilor statistice care pot lua mai mult de două niveluri. În statistică, dacă nu se face nici o altă precizare, atunci când se vorbeşte de caracteristică, în general se înţelege o caracteristică nealternativă.

d) După *modul lor de exprimare*, caracteristicile se pot împărţi în două grupe principale: *numerice* (cantitative) şi *atributive* (calitative).

Caracteristicile numerice reprezintă însuşiri ale unităţilor, care prin însăşi natura lor, sunt măsurabile şi se pot exprima cifric, numeric sau cantitativ, de exemplu: valoarea producţiei, numărul scriptic al personalului, vârsta, vechimea în producţie, productivitatea muncii etc.

Caracteristicile atributive sunt însuşiri ale unităţilor statistice care se exprimă prin cuvinte, de exemplu: starea civilă a populaţiei, ramurile economiei naţionale, sectoarele producţiei sociale etc. Atunci când numărul variantelor caracteristicilor atributive este foarte mare, se întocmesc *nomenclatoare statistice* cu gradul de amănunţime cerut de cercetarea statistică care se întreprinde.

e) După *treapta cercetării* în care apar, caracteristicile se pot împărţi în două grupe: *primare* şi *derivate*.

Caracteristicile primare reprezintă acele însuşiri ale unităţilor ale căror valori rezultă direct din înregistrarea statistică a colectivităţilor şi ele se exprimă, de regulă în mărimi absolute sau prin cuvinte.

Caracteristicile derivate rezultă din prelucrarea statistică a informaţiilor obţinute prin înregistrarea statistică şi se pot exprima în mărimi absolute, în mărimi medii şi în mărimi relative.

În statistică noţiunea de caracteristică cantitativă se înlocuieşte uneori cu noţiunea de *variabilă*. Un anumit nivel al caracteristicii poartă denumirea de *variantă a caracteristicii* sau *variantă a unei anumite variabile*.

Scanned with OKEN Scanner

Din punct de vedere a *naturii variaţiei*, variabilele statistice se împart în două categorii: *discrete* (discontinue) şi *continue*.

Se spune că *o variabilă este discretă* dacă variantele ei posibile pe un anumit interval sunt *valori izolate*. Cazul cel mai frecvent de variabile discrete este acela în care valorile ei posibile sunt numere întregi.

În general, o variabilă aleatoare discretă este definită de repartiţia :

$$X = \begin{pmatrix} x_1, x_2, \ldots, x_i, \ldots, x_n \\ p_1, p_2, \ldots, p_i, \ldots, p_n \end{pmatrix}$$

unde : x_i = valorile posibile ale caracteristicii X, $i = \overline{1,n}$;

p_i = probabilitatea ca X să ia o valoare x_i notată obişnuit $P(X=x_i)= p_i$;

iar

$$\sum_{i=1}^{n} p_i = 1$$

Dimpotrivă, *o variabilă este continuă* dacă variantele ei posibile sunt în număr infinit în cadrul unui anumit interval. Astfel, vârsta persoanelor este o caracteristică cantitativă al cărei număr de variante poate fi considerat infinit în interiorul intervalului *0 - 100 ani*.

O variabilă aleatoare continuă este definită de funcţia ei de repartiţie

$$F(x)=P(X<x)$$

care are următoarele proprietăţi:

1) $F(-\infty) = 0$

2) $F(+\infty) = 1$

3) $F(x)$ *este monoton crescătoare*

4) $P(X\varepsilon[a,b] = \int_a^b dF(x)dx = \int_a^b f(x)dx$

Scanned with OKEN Scanner

Variantele unei caracteristici trebuie să fie în acelaşi timp *incompatibile* şi *exhaustive*, adică fiecare unitate statistică componentă a unei colectivităţi trebuie să poarte o variantă a unei caracteristici şi numai una.

În general, toate caracteristicile legate de spaţiu (lungime, lăţime, înălţime, suprafaţă, volum etc.), de timp (vârstă, vechime, perioadă etc.), de greutate (kilograme, tone) cu multiplii şi submultiplii lor sau din combinaţia acestor elemente (viteză, debit, densitate, gravitate etc.) sunt *variabile statistice continue*.

Distincţia teoretică între variabilele continue şi cele discrete în practică poate deveni *arbitrară*, deoarece toate mărimile rezultate din măsurări sunt discrete datorită impreciziei lor. În domeniul economic această distincţie este şi *convenţională* astfel, o variabilă care poate lua un număr mare de variante, chiar dacă aceste valori sunt mărimi izolate (discrete) în practică ca urmare a procesului de rotunjire, variabila respectivă va fi considerată drept continuă (de exemplu: salariile personalului din diferite întreprinderi şi instituţii, cifra de afaceri a unei colectivităţi de societăţi comerciale etc.).

- *Variabilele cantitative* pot fi:
  - ◊ *metrice*, cele care posedă un zero natural, o origine absolută şi se măsoară cu ajutorul unei *scări proporţionale* sau de *raport*;
  - ◊ *scalate*, nu posedă un zero natural, au origine arbitrară şi se măsoară cu ajutorul unei scări de intervale.

- *Variabilele calitative* sunt variabile statistice ale căror variante (modalităţi) reprezintă diferite rubrici ale unui nomenclator astfel stabilit încât acestea să fie incompatibile şi exhaustive. Variabilele susceptibile de a purta două variante se mai numesc *binare*, *booliene* sau *dichotomice*.

Variabilele calitative pot fi de două feluri:

- ◊ *ordinale*, când este posibil să fie clasificate după o dimensiune sau o direcţie a unei proprietăţi alese (considerate), în sensul că unitatea statistică posedă grade diferite ale proprietăţii considerate;
- ◊ *nominale*, când variantele lor nu pot fi decât numite, când rubricile sau codurile corespunzătoare ale nomenclatorului nu sunt decât etichete care permit identificarea lor în diverse stări posibile. Aceste modalităţi (variante) nu pot fi ordonate. Scala nominală este, deci, mai mult un procedeu de clasificare decât de măsurare, ea dă numai o relaţie de echivalenţă asupra cărora se pot face operaţii matematice simple.

Din punct de vedere al *modului cum reflectă realitatea* (după natura lor), variabilele statistice se pot împărţi în două categorii:

Scanned with OKEN Scanner

• *variabile certe* (precise), iau valori care sunt observate, adică rezultate dintr-un raţionament *a posteriori* în cadrul analizei retrospective;

• *variabile aleatoare*, iau valori care nu s-au observat (încă), fiind rezultate dintr-un raţionament *a priori* în cazul analizei prospective.

Din punct de vedere al *gradului (al scării) de măsurare*, variabilele statistice pot fi:

• *cantitative*:

◊ *variabile metrice (măsurabile)* ;

◊ *variabile scalate*.

• *calitative*:

◊ *variabile ordinale;*

◊ *variabile nominale*.

Pentru studierea variabilelor continue, grupele de variante apropiate ca mărime se pot înscrie în intervale, cărora li se precizează limitele şi numărul de unităţi statistice care se încadrează în ele (frecvenţa sau ponderea).

2.4. Cuantificarea fenomenelor sociale

Prin *cuantificarea fenomenelor sociale* se înţelege transpunerea în formă cifrică, numerică, a caracteristicilor acestor fenomene pentru a înlesni *comparabilitatea, prelucrarea, analiza, sinteza şi prognoza* lor.

Problema cuantificării caracteristicilor fenomenelor sociale reprezintă o problemă de bază a ştiinţelor sociale în general şi a statisticii în special, care i-a preocupat dintotdeauna pe cercetători, dar mai ales în epoca modernă, pe măsura creşterii exigenţelor faţă de determinările ştiinţifice. De aceea, studiile referitoare la această problemă, apărute în ultimile decenii, sunt tot mai numeroase.

După felul cum se pretează la cuantificare, fenomenele sociale se pot clasifica în două categorii: unele care sunt *prin natura lor măsurabile* şi altele care pot fi *măsurate cu aproximaţie*.

Fenomenele sociale *prin natura lor măsurabile* sunt numeroase în viaţa socială, cuprinzând în general fenomenele demografice, fenomenele economice, elementele geografice, diferite fenomene cu privire la viaţa politică şi culturală (tirajul ziarelor, revistelor şi cărţilor, numărul cititorilor din biblioteci etc.).

Fenomenele sociale *măsurabile cu aproximaţie* se referă în special la opiniile şi comportamentele colectivităţilor umane în probleme şi împrejurări determinate. Măsurarea opiniilor şi comportamentelor colectivităţilor umane nu poate fi efectuată

Scanned with OKEN Scanner

decât prin compararea intensităţilor sau calităţilor diferite sub care se manifestă acestea la diferite persoane. Pentru efectuarea unor comparaţii cifrice se stabilesc anumite scări de mărimi numite *scalograme*.

Scalograma este o reprezentare grafică cu ajutorul căreia se măsoară şi se ilustrează *intensitatea opiniilor* şi se prezintă sub forma unei axe verticale sau orizontale cuprinzând ranguri de tipul *cu totul de acord, de acord, nu sunt de loc de acord*, pe care cercetătorul situează opiniile persoanelor anchetate. Apoi fiecărui rang i se atribuie o valoare numerică (de exemplu -2, -1, 0, +1, +2).

Scanned with OKEN Scanner

II. OBTINEREA, SISTEMATIZAREA SI PREZENTAREA INFORMATIILOR STATISTICE

Scanned with OKEN Scanner

Scanned with OKEN Scanner

3. OBŢINEREA INFORMAŢIILOR STATISTICE

4. GRUPAREA INFORMAŢIILOR STATISTICE

5. PREZENTAREA INFORMAŢIILOR STATISTICE

Scanned with OKEN Scanner

3. OBŢINEREA INFORMAŢIILOR STATISTICE

Aşa cum am arătat anterior, procesul cercetării statistice a fenomenelor colective are un caracter complex şi se desfăşoară într-o serie de etape succesive (formularea ipotezelor sau a problemelor, obţinerea şi sistematizarea informaţiilor statistice, prelucrarea, analiza, sinteza şi prognoza lor), fiecare cu o anumită sferă de probleme.

Obţinerea informaţiilor statistice are o importanţă deosebită pentru buna desfăşurare şi reuşita întregii cercetări statistice. De modul cum este pregătită şi se desfăşoară această etapă a cercetării depind toate celelalte etape şi chiar rezultatul ei final. Dacă informaţiile culese în această etapă sunt incomplete sau inexacte, aceste deficienţe se perpetuează în cadrul prelucrării şi analizei ulterioare şi uneori se amplifică, denaturând cunoaşterea veridică a realităţii, ca şi buna desfăşurare a procesului de luare a deciziilor. Cunoaşterea problemelor şi particularităţilor acestei etape este indispensabilă pentru orice statistician - cercetător sau practician - în înţelegerea şi utilizarea informaţiilor statistice.

Informaţiile statistice care circulă prin diverse canale pot fi utilizate pentru verificarea empirică a unei teorii, pentru descrierea numerică a unui fenomen economic, demografic, social, pentru a lua o decizie de politică economică la diferite niveluri de agregare a economiei ca şi în viaţa cotidiană. De felul cum această informaţie este obţinută şi transmisă depinde percepţia realităţii la care se referă, oportunitatea deciziilor individuale şi a celor colective, de la întreprindere până la autorităţile publice centrale ale unui stat, ca şi pe plan internaţional.

În mod obişnuit se aprecizează că datele statistice poartă în ele însele o rigoare care exclude îndoiala, iar prezentarea informaţiilor ar fi o fază neutră faţă de activitatea statistică care îi garantează obiectivitatea. Experienţa statistică arată că această constatare nu este întru totul exactă. Informaţiile statistice provenite din observări totale sau parţiale nu pot fi exacte în sensul absolut al cuvântului, numeroase dificultăţi care

apar în procesul de culegere, verificare, cuantificare, prelucrare şi prezentare determină producerea unor erori atât de natură obiectivă, cât şi subiectivă.

În prezent problemele obţinerii, sistematizării şi prezentării datelor statistice sunt stimulate şi uşurate de utilizarea calculatoarelor electronice. De aceea este necesar să fie cunoscute principiile fundamentale de prezentare a informaţiilor statistice, nu numai pentru construirea pe calculator a formei celei mai adecvate (cantităţii şi calităţii informaţiilor) a tabelelor şi graficelor, ci şi pentru utilizarea şi interpretarea corectă a informaţiilor statistice provenite din diferite surse.

3.1. Înregistrarea statistică şi felurile ei

Pe măsura dezvoltării sale ca ştiinţă, statistica şi-a perfecţionat în mod continuu metodele şi tehnicile de obţinere a informaţiilor cu privire la fenomenele colective.

Ţinând seama de importanţa etapei obţinerii informaţiilor statistice în cadrul procesului de cunoaştere a fenomenelor de masă, ca şi de complexitatea problemelor ce urmează a fi rezolvate în această etapă, efectuarea ei cu succes nu poate fi concepută fără un *plan judicios* întocmit, care trebuie să se refere la *scopul înregistrării* materialului faptic, la *obiectul înregistrării*, la *unitatea de înregistrare*, la *programul înregistrării* şi la o serie de *probleme organizatorice* (personal, formulare etc.). Informaţiile obţinute în urma înregistrării fiecărei unităţi a colectivităţii cercetate constituie materialul statistic iniţial, *materia primă* a statisticii, pe baza căreia se înfăptuieşte cunoaşterea ştiinţifică.

În teoria statistică se cunosc şi în practică se utilizează numeroase tipuri şi forme de înregistrări care se pot sistematiza prin clasificare după criterii diferite: după gradul de cuprindere a colectivităţii generale, după timpul de înregistrare, după procedeul de înregistrare şi după alte criterii.

I. Din punct de vedere *al gradului de cuprindere a colectivităţilor* studiate, în procesul de înregistrare deosebim: *înregistrări totale* şi *înregistrări parţiale*.

Înregistrarea totală sau *exhaustivă* înseamnă consemnarea într-o formă unitară a tuturor unităţilor componente ale unei colectivităţi statistice cu caracteristicile lor, care interesează o anumită cercetare.

Scanned with OKEN Scanner

Înregistrarea parţială se efectuează asupra unei părţi din unităţile care compun colectivitatea cercetată, cu caracteristicile cuprinse în studiu, în aşa fel încât partea înregistrată să fie reprezentativă pentru ansamblul colectivităţii respective.

Indiferent de forma de înregistrare (totală sau parţială), informaţia obţinută trebuie să servească pentru caracterizarea colectivităţii generale. Înregistrarea statistică mai poartă uneori denumirea de *observare statistică.*

II. Din punct de vedere *al timpului de înregistrare* se disting: *înregistrări curente, înregistrări periodice* şi *înregistrări ocazionale.*

Înregistrarea curentă se efectuează cu caracter continuu, sistematic şi operativ asupra faptelor şi fenomenelor în timpul, la locul şi pe măsura apariţiei lor.

Înregistrarea curentă oferă posibilitatea urmăririi operative a îndeplinirii programelor de dezvoltare, a descoperirii rezervelor nefolosite şi ajută la conducerea operativă a unităţilor economice. Ea îmbracă, în general, forma *documentaţiei primare*, care stă la baza evidenţei tehnico-operative, a evidenţei contabile şi a celei statistice.

Înregistrările periodice ale colectivităţilor de stări se fac la anumite intervale de timp, dinainte stabilite, egale sau neegale. Ele mai poartă şi denumirea de *înregistrări la un moment dat*, deoarece prezintă starea fenomenelor colective la un anumit moment, instantaneu ca o fotografiere, spre deosebire de înregistrările curente care se aseamănă cu filmarea.

În statistică, înregistrările periodice îmbracă forma recensămintelor populaţiei, locuinţelor, animalelor, inventarierilor statistice, efectuate cu scopul evaluării şi reevaluării mijloacelor fixe din economie, inventarierii mijloacelor circulante etc.

Înregistrările ocazionale sau *unice* se utilizează, în cazul unor colectivităţi de fapte şi evenimente discontinue, ocazionale, neobişnuite, de exemplu, înregistrarea clădirilor efectuată în luna martie 1977 pentru localităţile din *România* care au fost afectate de distrugerile cauzate de cutremurul din 4 martie 1977. Se pot înregistra în mod ocazional şi anumite probleme de o importanţă deosebită, al căror studiu ar trebui aprofundat, cum ar fi, de exemplu, bugetul de timp al muncitorilor, al femeii ocupate în producţie, al cadrelor didactice din învăţământ, al studenţilor etc.

Aceste trei feluri de înregistrare se utilizează în mod combinat în statistică, iar în cadrul lor rolul principal îl are înregistrarea curentă, care serveşte ca bază documentară pentru întocmirea evidenţei statistice.

III. Din punct de vedere *al procedeului folosit* înregistrările pot fi de mai multe feluri: *înregistrări directe, înregistrări pe bază de documente, înregistrări pe bază de interogare* şi *autoînregistrări.* Forma organizatorică a înregistrării, ca şi perioadele de

Scanned with OKEN Scanner

înregistrare sunt determinate în primul rând de natura fenomenelor sociale supuse cercetării.

Înregistrarea directă reprezintă consemnarea nemijlocită a fenomenelor în documente şi se efectuează prin contactul direct al cercetătorului cu fiecare unitate a colectivităţii, determinându-i nivelurile de dezvoltare a caracteristicilor şi înregistrându-le în formulare (recensămintele populaţiei, recensămintele animalelor, inventarierile statistice, toată documentaţia primară, cercetările prin sondaj ale calităţii producţiei etc.).

Înregistrarea pe bază de documente se face pe baza unor documente oficiale în care au fost consemnate fenomenele în timpul producerii lor, de exemplu, întocmirea evidenţei statistice pe baza evidenţei primare, a evidenţei tehnice operative şi a evidenţei contabile. Înregistrarea pe bază de documente înseamnă consemnarea datelor din documente anterioare şi transcrierea lor în formulare speciale. În condiţiile unei evidenţe primare bine organizate, înregistrarea pe bază de documente este tot atât de exactă ca şi înregistrarea directă, dar este mai puţin costisitoare.

Înregistrarea prin interogare se efectuează pe baza interogării de către cercetător a persoanelor supuse cercetării sau a celor care au fost de faţă la producerea fenomenului cercetat. Înregistrarea prin interogare se utilizează şi atunci când cercetătorul nu poate lua contact direct cu unităţile colectivităţii cercetate sau când fenomenele s-au produs anterior, fără a fi fost consemnate în documente, ca şi atunci când, deşi acesta are posibilitatea unui contact direct cu unităţile cercetate, nu este în măsură să stabilească singur nivelul de dezvoltare al anumitor caracteristici. Precizia informaţiilor obţinute dintr-o înregistrare prin interogare depinde de o serie de factori subiectivi, cum ar fi: exactitatea cu care faptele petrecute anterior s-au reflectat în conştiinţa celui interogat, corectitudinea punerii întrebărilor şi buna credinţă a recenzorului şi a celui recenzat.

Înregistrarea prin interogare se poate efectua şi prin *corespondenţă*.

Autoînregistrarea poate fi socotită o formă specială de înregistrare prin interogare, în care înregistrarea propriu-zisă se efectuează de către persoana cercetată, pe formulare speciale primite de la cercetător, de exemplu, anchetele statistice. În acest caz, cercetătorul înmânează sau trimite prin poştă formularele cu instrucţiunile de completare a lor celui interogat şi apoi le ridică după ce verifică în prealabil exactitatea acestora. Desigur, această formă de înregistrare a informaţiilor statistice este mai puţin costisitoare, dar şi rezultatele ei sunt mai puţin precise.

Scanned with OKEN Scanner

3.2. Verificarea autenticităţii informaţiilor statistice

În urma înregistrării informaţiilor statistice cu privire la fenomenele colective pot să apară *diferenţe* între nivelurile înregistrate ale unor anumite caracteristici şi nivelurile lor reale, *diferenţe* care poartă denumirea de *erori de înregistrare.*

Cauzele producerii erorilor de înregistrare sunt numeroase: lipsa de experienţă a personalului care efectuează înregistrarea, neglijenţa, insuficienta precizie a formularelor de înregistrare, ţinerea inexactă a evidenţelor, care stau la baza înregistrărilor statistice, erorile de măsurare etc. De aceea, se poate afirma că *nici o înregistrare statistică nu poate fi considerată exactă în sensul absolut al cuvântului.* Acest fapt a determinat să se manifeste o anumită neîncredere din partea nespecialiştilor faţă de informaţiile statistice, ca şi faţă de statistică în general. Aşa după cum arătau *G.U.Yule* şi *M.G. Kendall: "Atitudinea necunoscătorului faţă de statistică este admirabil rezumată în observaţia că omenirea este împărţită în două categorii şi anume, în acei care cred că cifrele pot exprima orice şi acei care cred că cifrele nu pot exprima nimic*"[204,p.23]. Aceste atitudini extreme cu privire la informaţiile statistice îşi au anumite explicaţii în statistică. Uneori asemenea date sunt utilizate, în mod nejustificat, ca argumente pentru anumite idei, alteori sunt folosite pentru a susţine o afirmaţie numai aparent adevărată. Puţine sunt metodele care au o arie aşa de largă de utilizare ca cele statistice şi nici o altă metodă nu necesită mai multă grijă în aplicarea ei. Statistica este una din acele ştiinţe ai cărei reprezentanţi trebuie să dea dovadă de o adevărată artă a discernământului, să privească întotdeauna cu o atitudine critică informaţiile statistice.

Este extrem de preţioasă în acest sens ideea lui *L.March* exprimată cu peste 70 de ani în urmă, potrivit căreia "*Critica este viaţa ştiinţei. Numai ea poate da valoare observaţiilor făcute, căci fără un control serios amestecul de adevăr şi fals ar împiedica orice lucrare ştiinţifică riguroasă, scepticismul ar triumfa şi nu ar mai fi în fapt ştiinţă adevărată*"[125,p.116].

Desigur, erorile de înregistrare influenţează cunoaşterea reală a colectivităţilor. De aceea, este necesar să fie cunoscute tipurile de erori care pot să apară cu ocazia înregistrărilor statistice, cauzele care determină apariţia lor, ce măsuri pot fi luate pentru a le preveni, cum pot fi descoperite şi înlăturate, pentru a asigura în final autenticitatea informaţiilor.

Scanned with OKEN Scanner

3.2.1. Clasificarea erorilor de înregistrare

După cum am văzut anterior, măsurările fenomenelor sociale şi ale celor naturale au un caracter aproximativ. Erorile de înregistrare sunt rezultatul consemnării nejuste sau neprecise a variantelor unei anumite caracteristici, cu ocazia înregistrării diferitelor unităţi ale colectivităţilor cercetate.

Erorile de înregistrare se pot datora fie intenţiei de denaturare a realităţii, de exagerare sau diminuare a proporţiilor unui anumit fenomen de către persoanele înregistrate sau chiar de ambele părţi şi, de aceea, ele poartă denumirea de *erori de înregistrare premeditate sau intenţionate*; fie neatenţiei, lipsei de memorie, înţelegerii greşite a formularelor şi instrucţiunilor de completare a lor şi în acest caz se numesc *erori de înregistrare nepremeditate sau neintenţionate*. Erorile de înregistrare premeditate, ca şi cele nepremeditate se pot referi numai la o parte sau la toate unităţile unei colectivităţi.

Erorile de înregistrare nepremeditate, după natura lor, se împart în două categorii: *întâmplătoare* şi *sistematice*.

Erorile întâmplătoare intră sub acţiunea legii numerelor mari şi nu influenţează în mod esenţial rezultatele medii, deoarece acţiunile lor într-un anumit sens sunt egal probabile cu acţiunile lor în sens contrar şi se compensează reciproc la nivelul colectivităţii. De exemplu, la efectuarea controlului statistic al calităţii producţiei, când se măsoară o anumită caracteristică, se comit erori de măsurare în plus sau în minus faţă de valorile reale ale caracteristicii diferitelor unităţi, dar pe ansamblu - dacă numărul variantelor este suficient de mare - abaterile într-un sens se compensează cu abaterile în sens contrar, iar valoarea medie va tinde spre valoarea reală a caracteristicii măsurate.

Erorile sistematice acţionează într-o anumită direcţie, bine determinată şi denaturează întotdeauna realitatea. Cele mai frecvente erori de înregistrare sistematice se produc cu ocazia înregistrărilor totale.

Erorile de înregistrare apar atât în cazul înregistrărilor totale, cât şi în cazul înregistrărilor parţiale. De regulă, ele vor fi mai mari în cazul înregistrărilor totale şi mai mici în cazul celor parţiale datorită, printre altele, şi volumului mai redus de muncă care determină un risc mai mic de comitere a erorilor. Aceasta nu înseamnă că informaţiile provenite dintr-o înregistrare parţială sunt mai exacte decât cele care provin dintr-o înregistrare totală, datorită faptului că primele, sunt afectate şi de un tip special de erori - *erori de reprezentativitate*.

Erorile de reprezentativitate se exprimă ca diferenţe între mărimea unui indicator sintetic (nivelul mediu al unei variabile, gradul de împrăştiere al variantelor

Scanned with OKEN Scanner

individuale în jurul nivelului ei mediu, diferiţi indicatori de structură etc.) obţinut pe baza eşantionului şi mărimea aceluiaşi indicator determinată pe baza unei înregistrări totale, cu condiţia ca eroarea de înregistrare să fie la fel de mare în ambele cazuri. Aceste erori sunt determinate de imposibilitatea reproducerii exacte în cadrul eşantionului a structurii colectivităţii generale.

3.2.2. Prevenirea şi controlul statistic al erorilor

Necesitatea obţinerii unor informaţii autentice a determinat dezvoltarea unor *metode şi tehnici variate de prevenire şi control al erorilor* care se comit cu ocazia înregistrărilor statistice. Datorită naturii lor, erorile premeditate pot fi prevenite, se pot chiar înlătura, prin convingerea persoanelor supuse înregistrării ca şi a celor care efectuează această operaţiune, asupra caracterului ei ştiinţific şi aplicativ general.

Pentru a se asigura autenticitatea datelor statistice este necesar *să se verifice în mod sistematic*, în toate etapele procesului de înregistrare, modul cum au fost înţelese programul şi formularele de înregistrare, instrucţiunile de completare a lor cu scopul de a preveni producerea erorilor.

Prevenirea producerii erorilor de înregistrare se poate face pe căi diferite, de exemplu prin:

- *testarea formularelor de înregistrare* prin anchete-test, recensăminte de probă etc.;
- *alegerea judicioasă a personalului de înregistrare*, mai ales pentru înregistrările totale;
- *instruirea personalului* care face înregistrarea şi, pentru unele cercetări speciale, chiar testarea aptitudinilor acestuia;
- *popularizarea lucrărilor statistice* şi formarea convingerilor în rândul populaţiei cu privire la însemnătatea ştiinţifică şi aplicativă a recensământului populaţiei, animalelor, pomilor, precum şi a diferitelor anchete statistice;
- efectuarea unui *control riguros asupra activităţii personalului* de înregistrare;
- *verificarea modului cum sunt ţinute evidenţele* ce stau la baza informaţiilor statistice.

Erorile nepremeditate pot fi mai greu prevenite şi de aceea datele statistice, înainte de a fi utilizate ca atare, trebuie analizate, controlate şi corectate.

Controlul datelor statistice este o acţiune ulterioară înregistrării şi se efectuează cu scopul de a asigura volumul complet şi calitatea datelor înregistrate. Acest control se poate efectua fie pe locul înregistrării, în mod operativ, fie la diferite organe

Scanned with OKEN Scanner

teritoriale sau centrale *(Direcţiile Judeţene de Statistică, Comisia Naţională pentru Statistică).*

Independent de locul unde se efectuează, în primul rând se procedează la un *control cantitativ* al datelor obţinute prin înregistrare, pentru a se constata dacă au fost cuprinse toate unităţile colectivităţii cercetate şi dacă la fiecare unitate au fost înscrise toate caracteristicile cuprinse în formulare. Astfel, în cazul diferitelor documente statistice se urmăreşte dacă s-au completat toate rândurile şi coloanele. Prin control cantitativ se verifică dacă este complet volumul informaţiilor statistice.

Pe lângă controlul cantitativ se efectuează în practica statistică şi un *control calitativ* cu ajutorul căruia se verifică înregistrarea corectă a indicatorilor stabiliţi. Controlul calitativ se efectuează sub formă de *control logic* şi sub formă de *control aritmetic.*

Controlul logic constă în a se verifica dacă legătura dintre diferite caracteristici înregistrate corespunde unei legături reale. Compararea acestor legături se poate face în timp, în spaţiu sau din punct de vedere calitativ.

Controlul aritmetic constă în verificarea calculelor din care au rezultat totalurile înregistrate.

Modalităţile de control prezentate aici nu epuizează gama posibilităţilor date de natura informaţiilor, de gradul de precizie urmărit, de nivelul de pregătire profesională şi de informare a statisticianului ca şi de legătura sa cu activitatea practică.

Scanned with OKEN Scanner

4. GRUPAREA INFORMAŢIILOR STATISTICE

4.1. Noţiunea de grupare a unităţilor statistice

Oamenii de ştiinţă aparţinând unor specialităţi diferite au recurs la gruparea sau clasificarea evenimentelor şi fenomenelor pe care le-au studiat încă din antichitate, datorită mai ales posibilităţilor limitate de memorizare a unui număr mare de noi fapte şi cunoştinţe, care s-au aflat, în toate etapele istorice ale dezvoltării umanităţii, într-un proces de creştere continuă. Astfel, apariţia şi dezvoltarea metodei grupărilor este strâns legată de necesitatea de a ordona, ierarhiza şi sistematiza fenomenele şi faptele din natură şi societate după criterii logice.

La început savanţii au recurs la *clasificări taxonomice*[1] care necesitau un criteriu uşor de aplicat, cum ar fi succesiunea cronologică a faptelor şi evenimentelor. Problemele teoretice ale clasificărilor se pare că au fost abordate pentru prima dată în *Grecia* antică de către *Platon (427 î.e.n - 348 î.e.n.)*, care a susţinut că *dihotomia*, diviziunea cu doi membri, este principiul raţional al oricărei clasificări. Acest principiu a fost ulterior combătut pe bună dreptate de către *Aristotel (384 î.e.n. - 322 î.e.n.)*, cel mai important discipol al lui *Platon*. Aristotel a arătat că în realitate, de cele mai multe ori, *clasificarea dihotomică este fie superficială, fie chiar imposibil de aplicat.*

Cu toată dezvoltarea metodei grupărilor din antichitate şi până în prezent, problemele fundamentale ale grupărilor nu sunt încă pe deplin rezolvate şi continuă să-i preocupe pe reprezentanţii tuturor ştiinţelor particulare în general şi cu deosebire pe logicieni şi statisticieni.

Statistica, studiind aspectele măsurabile ale fenomenelor colective, este în cea mai mare măsură interesată de clarificarea noţiunilor, principiilor şi metodelor de grupare a unităţilor unei colectivităţi, pentru a înlesni prelucrarea, analiza şi sinteza lor.

Informaţile statistice obţinute prin înregistrarea caracteristicilor unităţilor cercetate prezintă, în general, un număr mare de variante care nu permite formarea

---

[1] *Taxonomie* provine de la cuvintele greceşti *taxis* = aranjare + *nomos* = lege şi înseamnă ştiinţa legilor de clasificare.

Scanned with OKEN Scanner

unei imagini concrete şi complete asupra unor colectivităţi şi de aceea este necesară ordonarea unităţilor în funcţie de nivelul de dezvoltare al caracteristicilor esenţiale. Dacă numărul variantelor unei variabile statistice este foarte mare, aceasta nu dă posibilitatea evidenţierii anumitor regularităţi ce se manifestă în cadrul colectivităţii la diferite grupe de unităţi.

Împărţirea colectivităţilor statistice în grupe omogene de unităţi, după variaţia uneia sau a mai multor caracteristici esenţiale, comune tuturor unităţilor colectivităţii, poartă denumirea de grupare statistică, stratificare sau *clasificare*, iar *grupele* se mai numesc *straturi* sau *clase*.

Grupările au o importanţă deosebită în statistică deoarece, de modul cum sunt alcătuite grupele, depinde în bună măsură, cunoaşterea tipurilor calitative care se conturează în cadrul colectivităţilor social-economice, structura colectivităţilor, conţinutul indicatorilor ce urmează a fi calculaţi şi pe baza acestora însăşi rezultatele cercetării. De aceea, orice grupare a unităţilor unei colectivităţi trebuie să fie precedată de o analiză multilaterală a esenţei fenomenelor cercetate.

Pornind de la informaţiile sistematizate prin grupare, se pot obţine indicatori generalizatori, diferenţiaţi pe grupe, care să permită o analiză ştiinţifică a fenomenelor şi proceselor social-economice, a legăturilor esenţiale dintre acestea.

Caracteristica după a cărei variaţie se repartizează unităţile colectivităţii în grupe se numeşte *caracteristică de grupare*. Drept caracteristică de grupare se alege aceea care exprimă cel mai clar esenţa fenomenului studiat şi care răspunde cel mai bine scopului cercetării.

În procesul de grupare statistică este necesar să se ţină seama de următoarele reguli elementare:

1. Gruparea unităţilor statistice (respectiv a variantelor unei variabile) nu este raţională şi nici avantajoasă decât dacă numărul variantelor variabilei de grupare este suficient de mare.

2. Grupările nu prezintă interes decât în cazul în care numărul de grupe (clase) este relativ redus (de obicei până la 15).

3. Numărul de grupe trebuie să fie suficient de mare pentru a putea scoate în evidenţă anumite particularităţi ale distribuţiei unităţilor statistice în cadrul colectivităţii studiate.

4.2. Felurile grupărilor

Marea diversitate a grupărilor întâlnite în practică necesită clasificarea lor după diferite criterii, cum ar fi natura colectivităţilor şi scopul cercetării.

a. După *numărul de caracteristici* care stau la baza grupărilor se disting: *grupări simple* şi *grupări combinate* sau *complexe*.

Scanned with OKEN Scanner

Grupările simple se efectuează după o singură caracteristică de grupare, de exemplu, gruparea întreprinderilor din industrie după numărul salariaţilor, în anul 1990 (tabelul 4.1).

Regrupând întreprinderile industriale pe intervale de variaţie mai mari (vezi ultima coloană a tabelului 4.1) se poate obţine o imagine a *tipurilor calitative* care se diferenţiază în cadrul colectivităţii. Astfel, se disting tipurile:

- *întreprinderi mici*, care cuprind până la 200 de salariaţi;
- *întreprinderi mijlocii*, având numărul salariaţilor între 201 şi 1000;
- *întreprinderi mari*, în care lucrează între 1001 şi 5000 salariaţi;
- *întreprinderi foarte mari*, care au numărul salariaţilor mai mare de 5000.

Tabelul 4.1. Distribuţia întreprinderilor industriale după numărul salariaţilor la sfârşitul anului 1990

Grupe de întreprinderi după numărul salariaţilor	Numărul întreprinderilor		Numărul de salariaţi		
	Valori absolute	%	Valori absolute	%	
- 200	169	7,6	24288	1,9	1,9
201- 500	456	20,3	162505	5,2	15,9
501-1000	538	24,0	391386	10,7	
1001-2000	515	23,0	723116	18,9	53,8
2001-3000	245	10,9	592538	14,3	
3001-5000	197	8,8	768760	20,6	
5001-	121	5,4	1039263	28,4	28,4
TOTAL	2241	100,0	3701856	100,0	100,0

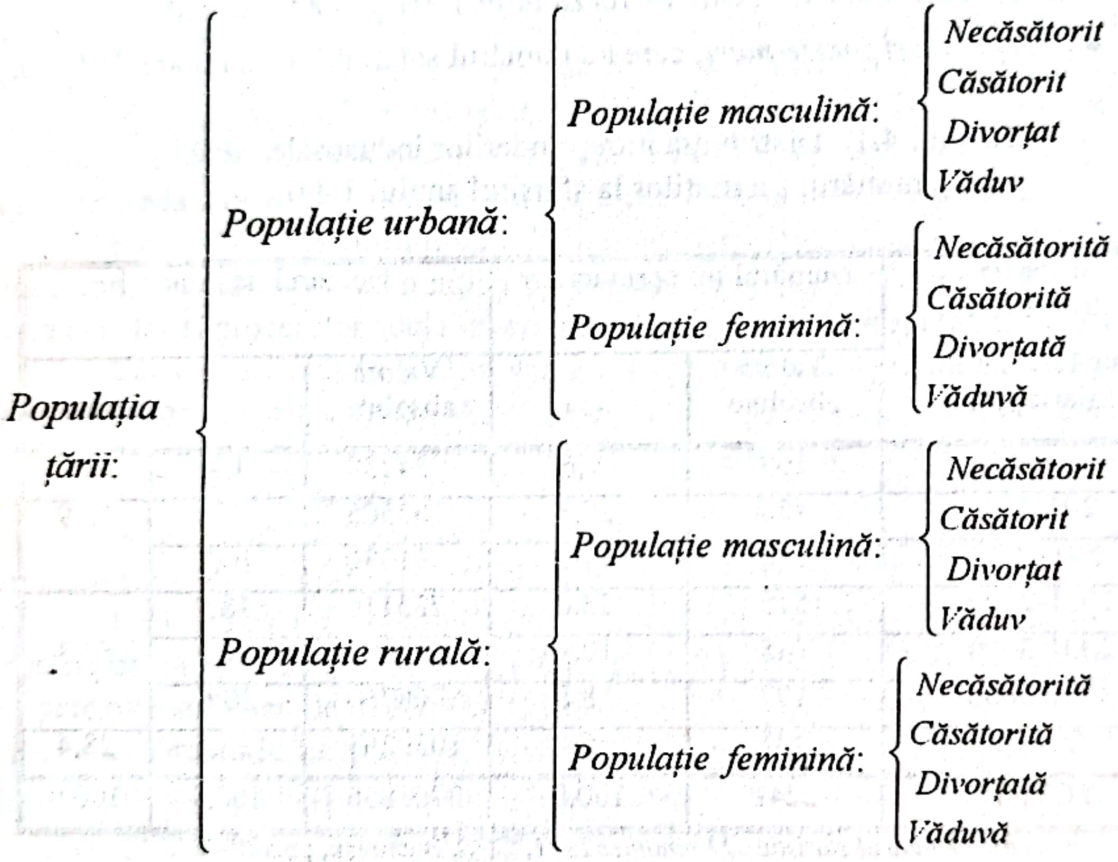
Sursa: *Anuarul statistic al României 1991*, CNS, Bucureşti, pp.*420-421*.

Grupările simple sunt utile pentru caracterizarea structurii colectivităţilor, pentru determinarea tendinţei de concentrare a fenomenelor, pentru evidenţierea tipurilor calitative care se conturează în dezvoltarea fenomenelor şi pentru rezolvarea altor probleme ale cercetării statistice, dar nu permit studierea legăturilor existente între fenomene. Acest neajuns se înlătură cu ajutorul grupărilor combinate.

Grupările combinate se întocmesc după două sau mai multe caracteristici, de exemplu, gruparea întreprinderilor industriale după numărul de salariaţi şi după mărimea fondurilor fixe, gruparea populaţiei după vârste, sexe şi medii sociale,

Scanned with OKEN Scanner

gruparea muncitorilor după vechimea în producţie şi după nivelul productivităţii muncii, gruparea studenţilor unei facultăţi pe secţii şi după rezultatele obţinute la învăţătură etc. În cadrul procesului de grupare combinată, colectivitatea se împarte în grupe după o anumită caracteristică, apoi aceste grupe se împart în alte grupe mai mici după o altă carcteristică şi aşa mai departe. Deci numărul grupelor creşte o dată cu numărul caracteristicilor de grupare considerate.



Schema 4.1. Gruparea combinată a populaţiei

În practica statistică, pentru a evita fărâmiţarea exagerată a colectivităţilor se aleg cel mult trei sau patru caracteristici de grupare, iar pentru fiecare caracteristică (de grupare) nu se formează un număr prea mare de grupe. Grupările

combinate se utilizează, în general, pentru cercetarea fenomenelor complexe între care există o legătură reciprocă. De exemplu, populaţia unei ţări înregistrată cu ocazia unui recensământ, poate fi grupată în mod combinat astfel:

- după mediul social în care trăieşte, se poate împărţi în populaţia urbană şi populaţia rurală;
- atât populaţia urbană, cât şi cea rurală se poate împărţi pe sexe în populaţia masculină şi populaţia feminină;
- la rândul lor populaţia masculină şi cea feminină, astfel obţinute, se poate împărţi în funcţie de caracteristica *stare civilă* în patru grupe mai omogene: necăsătorit, căsătorit, divorţat şi văduv.

Binenţeles că gruparea ar putea continua, în acest caz, mai departe după alte caracteristici sociale şi demografice. Modul de grupare combinată a populaţiei unei ţări se prezintă în mod sugestiv cu ajutorul schemei 4.1.

b. În funcţie de *conţinutul caracteristicilor de grupare* se disting următoarele feluri de grupări: *după caracteristici de timp, după caracteristici de spaţiu* şi *după caracteristici calitative.*

Grupările statistice după caracteristici de timp permit efectuarea analizei dependenţei fenomenelor cercetate de factorul timp şi evidenţierea tendinţei lor centrale de dezvoltare de lungă durată, determinarea şi separarea oscilaţiilor sezoniere şi a celor ciclice, pentru fenomenele care poartă acest caracter şi constituie informaţia de bază pentru efectuarea prognozelor. În cazul grupărilor după o caracteristică de timp, unităţile colectivităţii se repartizează în grupe după timpul în care au apărut, au existat sau au suferit un anumit eveniment.

Grupările după caracteristici de spaţiu se utilizează în practică pentru caracterizarea dezvoltării fenomenelor în profil teritorial, de exemplu: gruparea populaţiei pe judeţe, numărul mediu al salariaţilor pe judeţe, volumul comerţului exterior pe ţări, populaţia lumii pe ţări, producţia principalelor produse agricole pe ţări etc.

Grupările după o caracteristică calitativă se aplică în practica statistică pentru studierea structurii diferitelor colectivităţi şi uneori a tipurilor calitative care se separă în cadrul acestora, iar în acest caz se numesc *grupări tipologice.* De exemplu, gruparea întreprinderilor industriale după numărul salariaţilor, structura populaţiei pe sexe, medii, grupe de vârstă, categorii de localităţi, structura capitalului fix pe ramuri ale economiei naţionale etc.

c. După *forma de exprimare a caracteristicii* de diferenţiere, grupările se pot împărţi în două categorii:

Scanned with OKEN Scanner

• *grupări după caracteristici calitative*, exprimate prin cuvinte, numite şi *grupări tipologice* sau *clasificări*;

• *grupări după caracteristici cantitative* exprimate numeric.

d. La rândul lor grupările *după caracteristici cantitative* pot fi de două feluri, *după cum variaţia lor este discretă sau continuă.* În cazul unor *caracteristici discrete* gruparea se poate face, de regulă, pe variante ale acesteia, iar în cazul unor *caracteristici continue, pe intervale de variaţie.*

4.3. Stabilirea intervalelor de grupare

Uneori dezvoltarea fenomenelor sub aspectul nivelului unor caracteristici prezintă un număr foarte mare de variante, iar gruparea unităţilor după variante ar duce la fărâmiţarea exagerată a colectivităţii, într-un număr foarte mare de grupe mici, fără nici o semnificaţie. Acest inconvenient se înlătură dacă unităţile colectivităţii se reunesc în grupe pe anumite *intervale de variaţie*. Fiecare interval are o *limită inferioară* şi o *limită superioară*. În funcţie de natura colectivităţilor şi de scopul urmărit în cercetare, se pot stabili *intervale egale sau neegale de grupare.*

În practică *numărul de intervale de grupare*, considerat pentru o anumită variabilă continuă, depinde de mărimea preciziei măsurătorilor, de natura şi volumul colectivităţii studiate şi de scopul cercetării. Stabilirea unui număr mare de intervale poate determina apariţia unor *neregularităţi accidentale*, datorită numărului redus de unităţi cuprinse în fiecare interval, dimpotrivă, considerarea unui număr foarte mic de intervale poate determina o *pierdere de informaţie*. Între aceste două situaţii extreme de *precizie iluzorie şi costisitoare* - dată de un număr foarte mare de intervale mici - şi *informaţia insuficientă* există în mod logic un număr optim de intervale. Stabilirea numărului de grupe este rezultatul unui compromis care depinde, mai ales, de natura colectivităţii cercetate. În funcţie de natura caracteristicii studiate, numărul optim de intervale se situează în general între 10 şi 15.

Un interval de grupare $(x_{i-1} - x_i) \quad i = \overline{1,k}$ este descris de trei elemente:

• mărimea sau amplitudinea intervalului de grupare:

$$a_{x_i} = x_i - x_{i-1}$$

Scanned with OKEN Scanner

• centrul intervalului de grupare:

$$x_i' = \frac{x_{i-1} + x_i}{2}$$

• densitatea intervalului de grupare:

$$d_i = \frac{n_i}{a_{x_i}}$$

În cazul intervalelor neegale, d_i permite mai buna apreciere a greutăţii specifice (a ponderii relative) a unităţilor statistice care îi aparţin faţă de volumul total al colectivităţii.

Mărimea intervalelor este determinată de necesitatea obţinerii de efective comparabile - adică de acelaşi ordin de mărime - de la un interval la altul. În practică se folosesc atât intervale egale ca mărime, cât şi intervale neegale.

Limitele intervalelor trebuie să fie întotdeauna valori simple, de regulă numere întregi sau rotunjite.

În cazul unor intervale egale de grupare se recomandă, în literatura de specialitate, determinarea acestora după formula aproximativă a lui *H.A.Sturges*[2]:

$$d = \frac{x_{max} - x_{min}}{1 + 3{,}322 \log N} \qquad (4.1)$$

în care:

d = mărimea intervalului de grupare;

x_{max} = varianta maximă a caracteristicii de grupare;

x_{min} = varianta minimă a caracteristicii de grupare;

N = numărul unităţilor colectivităţii cercetate.

Cu ajutorul acestei formule se stabileşte, în funcţie de numărul unităţilor colectivităţii cercetate N, mărimea intervalelor d în care se pot grupa variantele

[2] H.A.Sturges: *The Choise of a Class Interval*, Journal of the American Statistical Association, vol. 21, 1926, p. 65-66

Scanned with OKEN Scanner

unei caracteristici pentru a limita volumul calculelor şi a nu pierde din informaţie. Numitorul formulei (4.1) reprezintă tocmai numărul intervalelor de grupare k .

Dacă se cunoaşte numărul de intervale egale de grupare k pentru o anumită colectivitate, atunci mărimea acestora d se poate determina cu ajutorul formulei:

$$d = \frac{x_{max} - x_{min}}{k} \tag{4.2}$$

Din relaţia (4.2) rezultă că, atunci când se cunoaşte mărimea intervalelor de grupare d , numărul acestora k se va determina astfel:

$$k = \frac{x_{max} - x_{min}}{d} \tag{4.3}$$

Trebuie să precizăm că asemenea formule de stabilire a mărimii intervalelor de grupare, cum este cea elaborată de statisticianul american *H. Sturges*, au mai mult o valoare orientativă şi nu pot, în nici un caz, să suplinească analiza calitativă a fenomenului cercetat. Aplicarea în mod mecanic a acestei formule poate să ducă la rezultate nesatisfăcătoare în practică. De regulă, pentru fenomenele în curs de formare şi dezvoltare, se recomandă a se utiliza *intervale neegale* de grupare, mai mici către partea inferioară şi mai mari spre partea superioară.

O altă problemă a grupărilor pe intervale este *stabilirea limitelor intervalelor de grupare.*

Se disting mai multe cazuri:

a. Când limita superioară face parte din interval (tabelul 4.2. sau tabelul 4.3.).

Scanned with OKEN Scanner

Tabelul 4.2.[x)]

Caracteristica	Numărul unităţilor
0 - 5	1
5 - 10	4
10 - 15	6
15 - 20	4
20 - 25	1
Total	16

[x)] *Date convenţionale*

Tabelul 4.3.[x)]

Caracteristica	Numărul unităţilor
0	
	1
5	
	4
10	
	6
15	
	4
20	
	1
25	
Total	16

[x)] *Date convenţionale*

b. Când se precizează în mod explicit limitele intervalelor de grupare:

Tabelul 4.4.[x)]

Caracteristica	Numărul unităţilor
0,1 - 5,0	1
5,1 - 10,0	4
10,1 - 15,0	6
15,1 - 20,0	4
20,1 - 25,0	1
T otal	16

[x)] *Date convenţionale*

Desigur, cea mai precisă modalitate de stabilire a limitelor intervalelor este varianta *b*, dar ea nu se utilizează în mod curent în practică, deoarece este relativ complicată şi de aceea se recurge în mod obişnuit la una din formele variantei *a*.

În activitatea statistică se întâlnesc şi situaţii în care nu se cunoaşte una dintre limitele intervalelor marginale (limita inferioară a primului interval sau limita superioară a ultimului interval). Dacă în cercetare este absolut necesară cunoaşterea acestor limite, se va recurge la obţinerea informaţiilor suplimentare cu privire la colectivitatea cercetată sau se vor utiliza mijloace indirecte de estimare, numai în cazul în care prima modalitate este inaplicabilă.

Scanned with OKEN Scanner

5. PREZENTAREA INFORMAŢIEI STATISTICE

5.1. Serii de repartiţie unidimensionale

5.1.1. Noţiunea, importanţa şi felurile lor

Datele statistice iniţiale, obţinute în urma observării fenomenelor şi proceselor colective, se prezintă sub formă brută ca o masă neordonată de date. Înainte ca acestea să fie supuse prelucrării, analizei ulterioare şi generalizărilor, ele trebuie să fie ordonate, grupate şi sistematizate, astfel încât să permită obţinerea unei imagini clare asupra colectivităţii la care se referă şi să se preteze la calcule algebrice ulterioare.

Ca rezultat al grupării şi centralizării datele statistice se sistematizează cu ajutorul unor tabele, sub forma a două şiruri paralele de date, dintre care unul reprezintă *variantele* sau *intervalele de variaţie* ale unei caracteristici (variabile), iar celălalt, numărul de unităţi statistice corespunzătoare fiecărei variante (sau fiecărui interval de variaţie), numit *frecvenţă*.

Să considerăm, pentru exemplificare, o colectivitate de fenomene constituită din n unităţi statistice pentru care se înregistrează un anumit număr de caracteristici. Pentru fiecare caracteristică X, Y, ..., Z se pot înregistra k niveluri distincte $(x_1, x_2, ..., x_i, ..., x_k), ..., (z_1, z_2, ..., z_i, ..., z_k)$, a căror succesiune poate fi considerată în ordinea apariţiei lor, în ordinea mărimii lor sau la întâmplare (fără un criteriu bine determinat). Un anumit nivel al caracteristicii (o anumită variantă a unei variabile) x_i se poate întâlni fie la o singură unitate statistică, fie la mai multe unităţi. Numărul unităţilor statistice n_i care poartă un anumit nivel x_i al caracteristicii se numeşte *frecvenţă absolută*, deoarece se exprimă în mărimi absolute.

Dacă se împarte fiecare frecvenţă absolută n_i la volumul total al colectivităţii, se obţine *frecvenţa relativă*, notată cu f_i, care se poate exprima sub formă de coeficienţi sau în procente (tabelul 5.1).

Prezentarea paralelă a celor două şiruri de date x_i şi n_i sau x_i şi f_i poartă denumirea de *serie de repartiţie* sau *distribuţie statistică* unidimensională (după o singură caracteristică) sau *distribuţie empirică* unidimensională. Seriile de repartiţie caracterizează fenomenele sub aspect static şi prezintă distribuţia

Scanned with OKEN Scanner

unităţilor unei colectivităţi în funcţie de variaţia unei variabile calitative de grupare, prezintă deci structura colectivităţilor.

Tabelul 5.1. Forma generală de prezentare a unei serii de repartiţie

Nivelurile caracteristicii x_i	Frecvenţa Absolută n_i	Frecvenţa relativă f_i
x_1	n_1	f_1
x_2	n_2	f_2
x_3	n_3	f_3
$\vdots$	$\vdots$	$\vdots$
x_i	n_i	f_i
$\vdots$	$\vdots$	$\vdots$
x_k	n_k	f_k
Total	$\sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} f_i$

Suma frecvenţelor absolute ale nivelurilor individuale este egală cu volumul colectivităţii la care se referă seria $\sum_{i=1}^{k} n_i = n$, iar suma frecvenţelor relative este egală cu unitatea:

$$\sum_{i=1}^{k} f_i = 1 \qquad \text{sau} \qquad \sum_{i=1}^{k} f_i \cdot 100 = 100$$

unde $f_i = \dfrac{n_i}{\sum_{i=1}^{k} n_i}$ şi $f_i \cdot 100 = \dfrac{n_i}{\sum_{i=1}^{k} n_i} \cdot 100$

Un anumit nivel al caracteristicii împreună cu frecvenţa lui reprezintă un *termen al seriei* (x_i, n_i) sau (x_i, f_i).

Scanned with OKEN Scanner

Dacă $n_1 = n_2 = n_3 = \cdots = n_i = \cdots = n_k = constant$, atunci se poate renunţa în practică la scrierea celui de al doilea şir n_i, iar şirul x_i poartă denumirea de *serie de repartiţie simplă.*

În general, frecvenţele n_i sunt diferite unele de altele (aceasta nu exclude egalitatea unora dintre ele) şi este necesar să se menţină cel de al doilea şir de valori, iar această prezentare paralelă a celor două şiruri de valori poartă denumirea de *serie de repartiţie cu frecvenţă.*

Datele statistice sistematizate cu ajutorul seriilor de repartiţie constituie un element esenţial al sistemului informaţional statistico-economic.

Alături de *distribuţiile empirice* unidimensionale există şi *distribuţii teoretice*, care corespund unei anumite legi de probabilitate cu o singură variabilă. Distribuţiile empirice iau naştere pe baza unor experienţe practice şi au un caracter statistic, de aceea se mai numesc şi *distribuţii statistice.*

Distribuţiile empirice ale fenomenelor reale din natură şi societate sunt studiate de către statistica aplicată, în timp ce distribuţiile teoretice sunt studiate de către calculul probabilităţilor şi statistica matematică.

Tabelul 5.2. Forma generală de prezentare a distribuţiilor frecvenţelor cumulate

x_i	n_i	Frecvenţa absolută cumulată $N(x_i)$	Frecvenţa relativă cumulată $F(x_i)$
x_1	n_1	$N(x_1)$	$F(x_1)$
x_2	n_2	$N(x_2)$	$F(x_2)$
⋮	⋮	⋮	⋮
x_{i-1}	n_{i-1}	$N(x_{i-1})$	$F(x_{i-1})$
x_i	n_i	$N(x_i)$	$F(x_i)$
x_{i+1}	n_{i+1}	$N(x_{i+1})$	$F(x_{i+1})$
⋮	⋮	⋮	⋮
x_{k-1}	n_{k-1}	$N(x_{k-1})$	$F(x_{k-1})$
x_k	n_k	$N(x_k)$	$F(x_k)$
Total	$\sum_{i=1}^{k} n_i$	-	-

Scanned with OKEN Scanner

Între distribuţiile teoretice şi cele empirice există o strânsă legătură. De multe ori este suficientă înlocuirea cuvântului *probabilitate* cu cel de *frecvenţă* pentru a se trece de la o demonstraţie privind o anumită proprietate a unei distribuţii teoretice la demonstraţia privind proprietatea analogă a unei distribuţii empirice.

Datele statistice, sistematizate cu ajutorul distribuţiilor empirice, constituie un element esenţial în cadrul cercetării statistice pe baza căruia se pot determina indicatorii derivaţi.

Frecvenţele empirice se pot cumula din aproape în aproape, ajungându-se astfel la stabilirea *distribuţiilor frecvenţelor cumulate.* Fie că este vorba de frecvenţa absolută sau de cea relativă, *frecvenţa cumulată* a unei valori empirice x_i a caracteristicii este egală cu suma frecvenţelor tuturor valorilor mai mici sau mai mari decât x_i, inclusiv aceea a lui x_i, în funcţie de sensul cumulării $N(x_i)$ şi $F(x_i)$ respectiv $N'(x_i)$ şi $F'(x_i)$ în cazul retrocumulării (tabelul 5.2).

Dacă cumularea se face pornind de la nivelul minim al lui x_i spre cel maxim, ultima frecvenţă cumulată este egală cu totalul frecvenţei seriei şi deci cu volumul colectivităţii:

$$N(x_k) = \sum_{i=1}^{k} n_i \qquad F(x_k) = \sum_{i=1}^{k} f_i$$

În mod analog, dacă cumularea se face începând de la nivelul maxim spre cel minim (*retrocumulare*), atunci frecvenţele cumulate corespunzătoare nivelului minim al caracteristicii $N'(x_1)$ şi $F'(x_1)$ vor fi egale cu volumul colectivităţii:

$$N'(x_i) = \sum_{j=k}^{k-i+1} n_j \qquad \text{sau} \qquad F'(x_i) = \sum_{j=k}^{k-i+1} f_j \qquad i = \overline{1,k}$$

Frecvenţele relative cumulate permit definirea *funcţiei cumulative a frecvenţelor* sau *funcţia de repartiţie* $F(x)$ astfel:

Scanned with OKEN Scanner

$$F(x)=\begin{cases}0 & dacă \quad x<x_1\\ f_1 & dacă \quad x_1\le x<x_2\\ \sum_1^2 f_i & dacă \quad x_2\le x<x_3\\ \cdots\cdots\cdots & \\ 1 & dacă \quad x\ge x_k\end{cases}$$

unde $x_1<x_2<x_3<\cdots<x_i<\cdots<x_k$

Aceasta înseamnă că funcţia de repartiţie empirică este nedescrescătoare şi variază între 0 şi 1 prin salturi succesive cu f_i.

Dacă cumularea se face de la x_k la x_1, *funcţia de repartiţie* $F'(x)$ va fi definită de relaţia:

$$F'(x)=\begin{cases}1 & dacă \quad x\le x_1\\ \sum_1^2 f_i & dacă \quad x_1<x\le x_2\\ \sum_2^3 f_i & dacă \quad x_2<x\le x_3\\ \cdots\cdots\cdots & \\ 0 & dacă \quad x>x_k\end{cases}$$

unde $x_1<x_2<x_3<\cdots<x_i<\cdots<x_k$

Rezultă că funcţia empirică retrocumulativă a frecvenţelor $F'(x)$ este descrescătoare şi variază între 1 şi 0 prin salturi succesive cu f_i.

Seriile de repartiţie au o mare răspândire în statistica social-economică, fiind utilizate pentru a reprezenta structura diferitelor fenomene sociale şi economice pe sfere ale economiei naţionale, pe sectoare de activitate, pe ramuri ale

Scanned with OKEN Scanner

economiei, pe forme de proprietate etc., de exemplu, distribuţia populaţiei după caracteristici demografice, sociale, economice etc.

După modul de exprimare a caracteristicii de grupare (numeric sau prin cuvinte), seriile de repartiţie se pot împărţi în două categorii:

- serii de repartiţie cu caracteristica exprimată numeric sau cantitativ, numite şi *serii cu atribut cantitativ*;
- serii de repartiţie cu caracteristica exprimată prin cuvinte, care se mai numesc şi *serii cu atribut calitativ*.

Dacă nivelurile individuale ale caracteristicii cantitative sunt numeroase, atunci ele se pot grupa pe anumite intervale de variaţie, ca în tabelul 5.3.

Tabelul 5.3. Serie de repartiţie cu caracteristica sub formă de intervale

Caracteristica (X)	Frecvenţa absolută (n_i)	Frecvenţa absolută cumulată (N_i)	Frecvenţa relativă (f_i)	Frecvenţa Relativă cumulată (F_i)
$x_0 - x_1$	n_1	N_1	f_1	F_1
$x_1 - x_2$	n_2	N_2	f_2	F_2
$x_2 - x_3$	n_3	N_3	f_3	F_3
⋮	⋮	⋮	⋮	⋮
$x_{i-1} - x_i$	n_i	N_i	f_i	F_i
⋮	⋮	⋮	⋮	⋮
$x_{k-1} - x_k$	n_k	N_K	f_k	F_k
Total	$\sum_{i=1}^{k} n_i$	-	$\sum_{i=1}^{k} f_i$	-

În cazul unor astfel de serii, frecvenţele (absolute sau relative) sunt funcţii de anumite intervale, care sunt comparabile de la un interval la altul, atunci când intervalele sunt egale şi necomparabile în cazul unor intervale neegale.

Pentru a deveni comparabile de la un interval la altul, şi în acest caz, se determină anumite mărimi abstracte, numite *frecvenţe transformate*, prin împărţirea frecvenţelor absolute sau a celor relative la mărimea intervalelor corespunzătoare, astfel:

Scanned with OKEN Scanner

$$n_i' = \frac{n_i}{\Delta_{x_i}} \qquad \text{şi} \qquad f_i' = \frac{f_i}{\Delta_{x_i}}$$

unde: n_i' - frecvenţa absolută transformată;

f_i' - frecvenţa relativă transformată;

$\Delta_{x_i} = x_i - x_{i-1}$

Frecvenţele transformate n_i' şi f_i' se mai numesc şi *frecvenţe unitare* sau *frecvenţe pe intervale unitare.*

Tabelul 5.4. Populaţia totală a *României* pe grupe de vârstă la recensământul din 7 ianuarie 1992

Grupa de vârstă (ani)	Numărul populaţiei	Grupa de vârstă (ani)	Numărul populaţiei
0 - 4	1 613 419	45 - 49	1 173 644
5 - 9	1 653 885	50 - 54	1 347 530
10 - 14	1 914 598	55 - 59	1 363 062
15 - 19	1 916 936	60 - 64	1 232 053
20 - 24	2 042 068	65 - 69	1 029 195
25 - 29	1 255 780	70 - 74	558 656
30 - 34	1 546 138	75 - 79	475 007
35 - 39	1 713 144	80 - 84	303 819
40 - 44	1 527 519	85 şi peste	143 582
Total			22 810 035

Sursa: *Anuarul statistic al României*, CNS, 1993, pp.96-98.

Termenii seriei iniţiale se pot grupa pe intervale egale sau neegale, în funcţie de natura fenomenului la care se referă şi de scopul ce se urmăreşte prin alcătuirea seriei.

Să considerăm acum seria de distribuţie cu caracteristica exprimată numeric, reprezentând gruparea municipiilor şi oraşelor din *România* după numărul populaţiei la recensământul din 7 ianuarie 1992 (tabelul 5.5).

Deci, din cele 260 de oraşe ale *României*, existente la data recensământului, un număr de 109 aveau o populaţie de peste 20 de mii de locuitori, iar 25 de oraşe aveau o populaţie de peste 100 de mii de locuitori, din care 12 peste 200 de mii de locuitori. Paralel cu dezvoltarea social-economică a ţării a avut loc o creştere continuă a gradului de urbanizare, care s-a concretizat în

Scanned with OKEN Scanner

sporirea numărului de locuitori în localităţile urbane vechi şi în cele nou înfiinţate, ca şi prin creşterea numărului de oraşe cu o populaţie mai numeroasă, cu importante implicaţii asupra structurii populaţiei pe medii.

Tabelul 5.5. Distribuţia municipiilor şi oraşelor din *România* după numărul populaţiei existente la recensământul din 7 ianuarie 1992

Grupe de oraşe după numărul de locuitori	Numărul municipiilor şi oraşelor	Numărul populaţiei	
		Persoane	%
sub 2000	1	1 969	*
2000 - 4999	10	37 848	0,3
5000 - 9999	54	414 173	3,3
10000 - 19999	86	1 183 395	9,6
20000 - 49999	61	1 910 006	15,4
50000 - 99999	23	1 723 713	13,9
100000 - 199999	13	1 827 016	14,8
200000 - 999999	11	3 226 154	26,0
1000000 şi peste	1	2 067 545	16,7
Total	260	12 391 819	100,0

* sub 0,1%

Sursa: *Anuarul statistic al României 1993*, CNS, Bucureşti, p.91.

Distribuţia municipiilor şi oraşelor din *România* s-a făcut pe intervale neegale de grupare, deoarece procesul de urbanizare în ţara noastră este în curs de desfăşurare. În procesul complex de transformare a localităţilor rurale în centre urbane, se diferenţiază o serie de etape cu o anumită succesiune şi semnificaţie. De aceea este necesar, pentru oraşele cu un număr redus de locuitori, să se formeze intervale de grupare mai mici, iar pe măsură ce numărul locuitorilor oraşelor creşte, va creşte şi mărimea intervalelor. În general, se poate aprecia că oraşele dintr-un anumit interval au trăsături şi roluri specifice în viaţa economico-socială din zonă.

Astfel, oraşele mici sunt fie centre industriale, comerciale sau agro-industriale nou înfiinţate, fie centre urbane mai vechi cu importanţă turistică restrânsă. Oraşele mari - municipiile - concentrează în ele o importantă activitate economico-socială şi culturală, şi sunt nuclee puternice de urbanizare pentru zonele din care fac parte.

Scanned with OKEN Scanner

Pentru ilustrarea seriilor de distribuţie cu atribut calitativ, considerăm datele din tabelul 5.6 în care se prezintă structura comerţului exterior al *României* pe grupe de mărfuri.

Frecvenţele sunt exprimate în procente, pentru fiecare grupă de mărfuri, faţă de total export şi respectiv total import.

Ca urmare a procesului de industrializare rapidă a *României*, în perioada 1950 - 1990 s-au produs mutaţii importante în structura comerţului exterior al ţării, în sensul creşterii exportului la maşini, utilaje, mijloace de transport, în general a produselor prelucrate şi scăderii în mod corespunzător a ponderii exportului de materii prime alimentare şi nealimentare.

Tabelul 5.6. Comerţul exterior al *României* pe grupe de mărfuri în anii 1950, 1960, 1970, 1980 şi 1990

Grupe de mărfuri	1950		1960		1970		1980		1990	
	Ex-port	Im-port	Ex-port	Im-port	Ex-port	Im-port	Ex-port	Im-port	Ex-port	Im-port
Maşini, utilaje şi mijloace de transport	4,2	38,3	16,17	33,6	22,8	40,3	24,9	24,6	30,8	23,7
Combustibil, materii prime minerale, metale şi materiale de construcţii	38,2	24,6	39,4	35,4	25,3	31,9	31,7	51,3	36,2	48,5
Produse chimice, îngrăşăminte, cauciuc	1,7	4,5	2,2	7,4	7,0	6,7	9,7	6,4	6,5	6,3
Materii prime alimentare şi produse prelucrate	28,9	21,4	14,7	13,4	10,2	10,1	4,8	5,7	4,2	6,6
Materii prime pentru producţia mărfurilor alimentare şi animale vii	11,6	0,7	9,4	2,6	4,5	2,3	4,2	5,8	0,4	5,0
Mărfuri alimentare	14,1	0,3	11,8	2,4	12,1	3,0	8,5	3,2	0,8	5,2
Mărfuri industriale de larg consum	1,3	10,2	5,8	5,2	18,1	5,5	16,2	3,0	21,1	4,7
Total	100	100	100	100	100	100	100	100	100	100
Indici cu baza 1950	100	100	338	266	871	805	40 ori	40ori	46 ori	30 ori

Sursa: Prelucrat după *Anuarul statistic al României*, CNS, 1991, p.600-601

Scanned with OKEN Scanner

5.1.2. Reprezentarea grafică a seriilor de repartiţie

Pentru a uşura analiza şi generalizarea rezultatelor, seriile statistice pot fi reprezentate grafic prin intermediul punctelor, liniilor, suprafeţelor şi a altor forme geometrice, prin figuri naturale şi simbolice, precum şi cu ajutorul hărţilor. Această metodă intuitivă, flexibilă şi foarte utilă pentru lămurirea, interpretarea şi analiza fenomenelor şi proceselor social-economice, poartă denumirea de *metoda reprezentării grafice a informaţiilor statistice.*

Reprezentarea grafică a seriilor de distribuţie urmăreşte să scoată în evidenţă diferite părţi ale colectivităţii studiate pe baza variaţiei caracteristicii şi a frecvenţelor corespunzătoare. Prin reprezentarea grafică a acestor serii se obţine o primă imagine asupra *formei distribuţiei*, asupra *structurii şi modificărilor structurale în timp şi spaţiu ale colectivităţii* şi se pot determina anumiţi *indicatori.*

Graficele utilizate pentru reprezentarea distribuţiilor statistice sunt numeroase. Forma şi tipul acestora depinde de natura fenomenului studiat şi de scopul urmărit. Astfel, în practică, se folosesc: diagrame prin benzi, histograme, diagrame triunghiulare, diagrame de structură, diagrame cu figuri naturale, simbolice etc.

Diagramele prin benzi sunt reprezentările grafice cele mai simple şi foarte răspândite în practică. Lungimea benzilor este direct proporţională cu indicatorii reprezentaţi, iar lăţimea lor este aceeaşi pentru toate benzile dintr-o diagramă. Aceste diagrame se utilizează în practică mai ales pentru reprezentarea grafică a seriilor de distribuţie cu caracteristica exprimată prin cuvinte (fig.5.1).

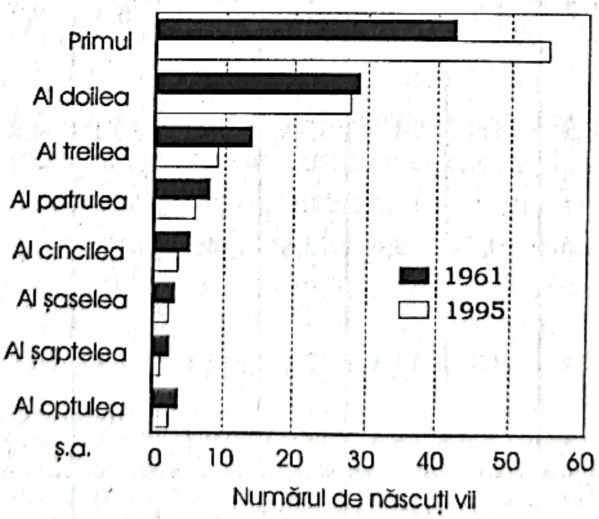


Figura 5.1. Distribuţia născuţilor-vii, după rangul lor, în anii 1961 şi 1995

Scanned with OKEN Scanner

Considerăm, pentru exemplificarea construcţiei diagramelor prin benzi, datele din tabelul 5.7. Din acest tabel se observă că numărul de născuţi-vii variază invers proporţional cu rangul lor.

În anul 1995 comparativ cu anul 1961, a avut loc o creştere a numărului născuţilor-vii de rangul unu şi o reducere corespunzătoare în cazul tuturor celorlalte ranguri.

Tabelul 5.7. Distribuţia născuţilor-vii din *România* în anii 1961 şi 1995 după rangul lor

Rangul născuţilor-vii	1961		1995	
	Născuţi-vii	%	Născuţi-vii	%
Primul	137 949	42,4	128742	54,4
Al doilea	91 539	28,2	65540	27,7
Al treilea	39 875	12,3	18670	7,9
Al patrulea	20 884	6,4	9092	3,8
Al cincilea	12 257	3,8	5255	2,2
Al şaselea	7 996	2,5	3584	1,5
Al şaptelea	5 526	1,7	2203	0,9
Al optulea şi peste	8 833	2,7	3554	1,5
Total	324859	100,0	236640	100,0

Sursa: *Anuarul statistic al României 1991*, CNS, p.72-74
Sursa: *Anuarul statistic al României 1995*, CNS, p.110-111

O formă specifică de diagramă prin coloane, utilizată pentru reprezentarea distribuţiilor statistice unidimensionale, este *histograma*.

Pentru construirea histogramei pe abscisă, se trece mărimea intervalelor de variaţie a caracteristicii, respectându-se principiul ca intervale egale să fie reprezentate prin distanţe egale. Pe ordonată se construieşte scara frecvenţelor. Pentru fiecare termen al seriei se ridică câte un dreptunghi având baza pe axa absciselor egală cu mărimea intervalului, iar înălţimea proporţională cu frecvenţa. Coloanele se construiesc alăturat, iar laturile superioare formează o linie frântă, în trepte, cu caracter de continuitate, care oferă o imagine clară asupra formei distribuţiei. Deoarece suprafeţele coloanelor sunt proporţionale cu frecvenţele reprezentate, suprafaţa totală a histogramei este proporţională cu volumul colectivităţii cercetate.

O problemă deosebit de importantă ce se ridică la construirea histogramelor este alegerea scărilor pe cele două axe de coordonate, deoarece de mărimea lor şi de raportul dintre ele depinde forma graficului. Rezolvarea acestei probleme se face pe baza experienţei fiecărui statistician, deoarece nu există criterii general valabile şi bine conturate. Se poate arăta totuşi că, cu cât mărimea

Scanned with OKEN Scanner

intervalelor pe axa absciselor este mai mare, cu atât variaţia fenomenului reprezentat apare mai uniformă. Dimpotrivă, intervalele prea mici fărâmiţează dezvoltarea fenomenelor evidenţiind şi variaţiile lor nesemnificative. Evident, între aceste două situaţii extreme există o variantă care este cea mai potrivită pentru fiecare fenomen reprezentat.

La construirea histogramelor se diferenţiază două cazuri distincte după cum seriile de repartiţie au intervalele egale sau neegale.

Pentru exemplificarea primului caz de construire a histogramei se consideră seria din tabelul 5.8.

Tabelul 5.8. Distribuţia unui grup de muncitori după nivelul productivităţii muncii[x)]

Numărul de piese lucrate într-o zi wh	Numărul muncitorilor n_i
210 - 215	3
215 - 220	5
220 - 225	10
225 - 230	20
230 - 235	35
235 - 240	50
240 - 245	60
245 - 250	30
250 - 255	25
255 - 260	10
260 - 265	8
265 - 270	4
Total	260

[x)] *Date convenţionale*

Histograma se poate construi cu ajutorul coloanelor ca în figura 5.2. Dacă în cazul histogramei prin coloane se omit liniile verticale care despart coloanele, histograma ia forma unei *diagrame în trepte* care scoate mai bine în evidenţă forma distribuţiei, conturul acesteia (fig.5.3).

Scanned with OKEN Scanner

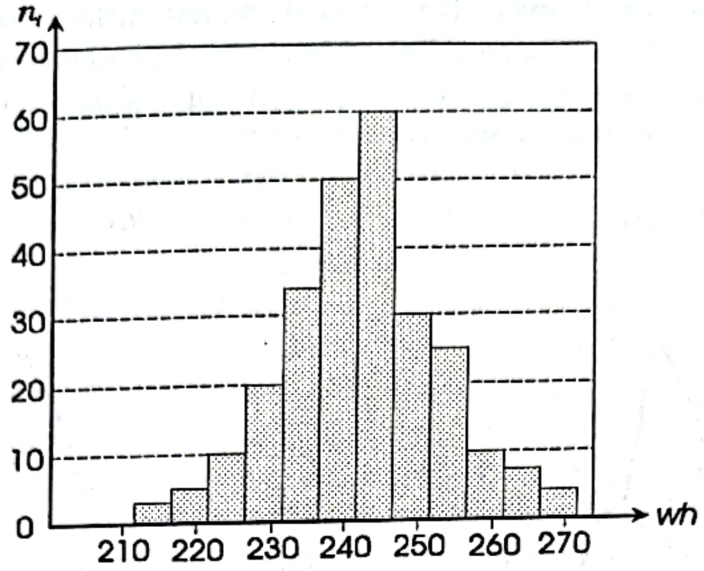


Figura 5.2. Histograma

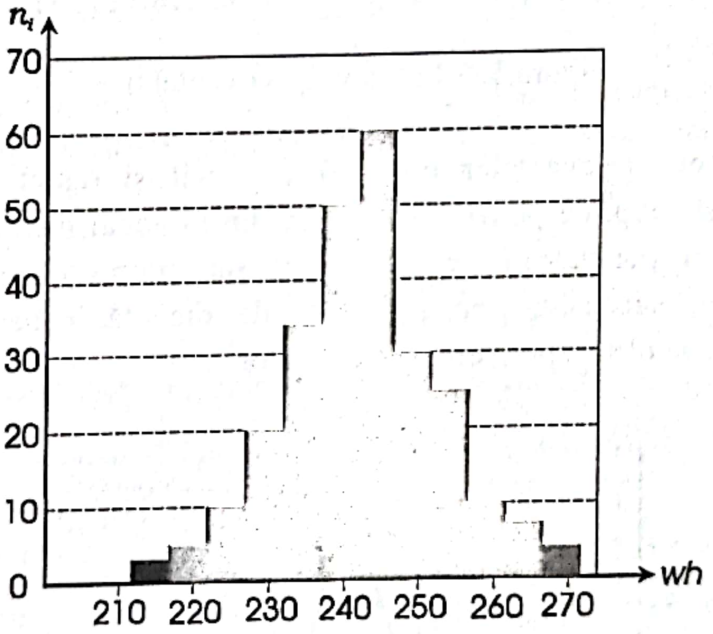


Figura 5.3. Diagrama în trepte

Poligonul frecvenţelor este un alt tip de grafic utilizat pentru reprezentarea seriilor de distribuţie cu atribut cantitativ. El se poate construi în două feluri:

a) Dacă se porneşte de la histograma seriei, atunci, pentru a obţine poligonul frecvenţelor, se unesc succesiv mijloacele capetelor de sus ale coloanelor histogramei prin segmente de dreaptă (fig.5.4a).

Scanned with OKEN Scanner

Aria poligonului frecvenţelor este egală cu aria histogramei pe baza căreia a fost construit (fig.5.4), deoarece linia frântă a poligonului frecvenţelor lasă în afară, din suprafaţa histogramei, o suprafaţă egală cu cea încorporată de poligonul frecvenţelor din exteriorul histogramei.

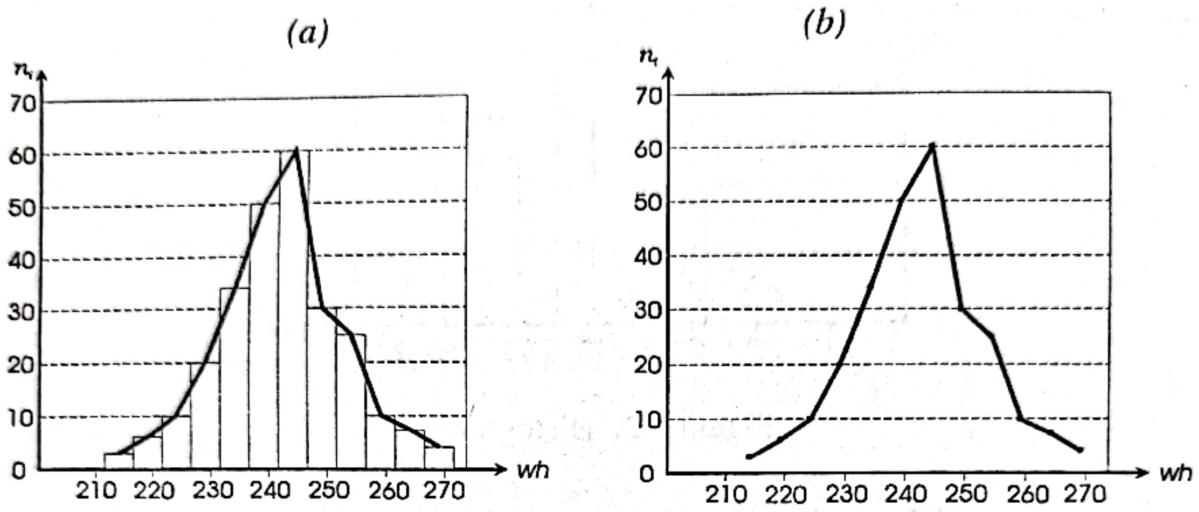


Figura 5.4. Poligonul frecvenţelor

b) Poligonul frecvenţelor poate fi construit şi direct, pe baza seriei (fig.5.4b). În acest scop, de pe axa absciselor, din mijlocul intervalelor, se ridică perpendiculare (linii punctate) ale căror înălţimi sunt proporţionale cu frecvenţele respective. Unind, succesiv, prin segmente de dreaptă, capetele de sus ale perpendicularelor, se obţine poligonul frecvenţelor

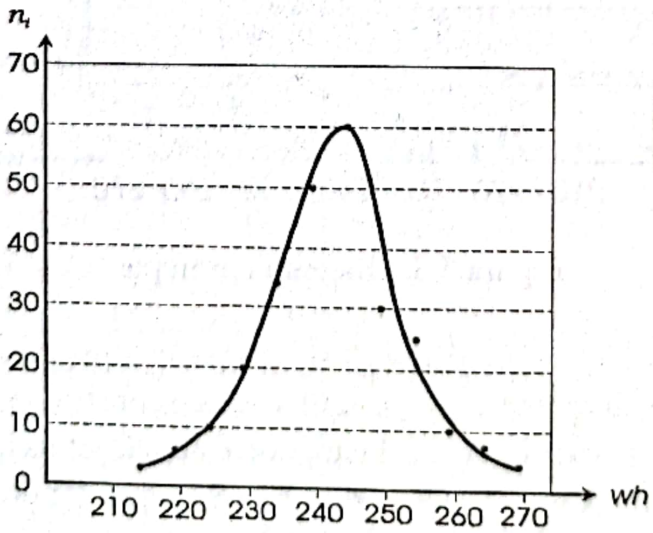


Figura 5.5. Curba frecvenţelor

Scanned with OKEN Scanner

Curba frecvenţelor sau *curba frecvenţelor ajustate vizual* se obţine prin înlocuirea liniei frânte, a poligonului frecvenţelor, cu o curbă continuă care va delimita aceeaşi suprafaţă ca şi poligonul frecvenţelor sau histograma şi care se va apropia în mare măsură de toate punctele empirice, fără a trece obligatoriu chiar prin ele (fig.5.5). Curba frecvenţelor reprezintă expresia geometrică a funcţiei teoretice care se apropie în cea mai mare măsură de punctele empirice, ea reprezintă o aproximare a legii teoretice de distribuţie.

În cazul distribuţiilor pe intervale neegale, dreptunghiurile care compun histograma trebuie construite în aşa fel încât, şi în acest caz, să se respecte principiul de bază al proporţionalităţii suprafeţei histogramei cu volumul colectivităţii. În acest scop, este necesar ca frecvenţele seriilor cu intervale neegale să fie transformate în frecvenţe corespunzătoare unor intervale egale.

Pentru exemplificarea construirii histogramei, în acest caz, să considerăm seria din tabelul 5.9.

Fiind o serie cu intervale neegale înainte de a-i construi histograma, este necesar ca frecvenţele n_i să fie transformate în frecvenţe pe intervale unitare n_i' (col.4 a tabelului 5.9) şi apoi multiplicate cu mărimea intervalului minim $(\Delta_{x_{min}})$.

Deoarece suprafeţele dreptunghiurilor, prin care reprezentăm grupele colectivităţii, trebuie să fie direct proporţionale cu frecvenţele, iar baza dreptunghiurilor de pe abscisă nu poate fi modificată, devine necesară recalcularea acestora după formula:

$$n_i'' = \frac{n_i}{\Delta_{x_i}} \cdot \Delta_{x_{min}}$$

Tabelul 5.9. Distribuţia muncitorilor după productivitatea zilnică a muncii

Numărul de piese lucrate într-o zi	Numărul lucrătorilor n_i	Mărimea intervalului Δ_{x_i}	$n_i' = \frac{n_i}{\Delta_{x_i}}$	$n_i'' = \frac{n_i}{\Delta_{x_i}} \cdot \Delta_{x_{min}}$
210-220	8	10	0,8	4,0
220-230	30	10	3,0	15,0
230-235	35	5	7,0	35,0
235-240	50	5	10,0	50,0
240-245	60	5	12,0	60,0
245-250	30	5	6,0	30,0
250-255	25	5	5,0	25,0
255-270	22	15	1,5	7,5
Total	260	-	45,3	226,5

Scanned with OKEN Scanner

Seria din tabelul 5.9 cu frecvenţele ei reale este reprezentată în figura 5.6, iar cu frecvenţele transformate în figura 5.7.

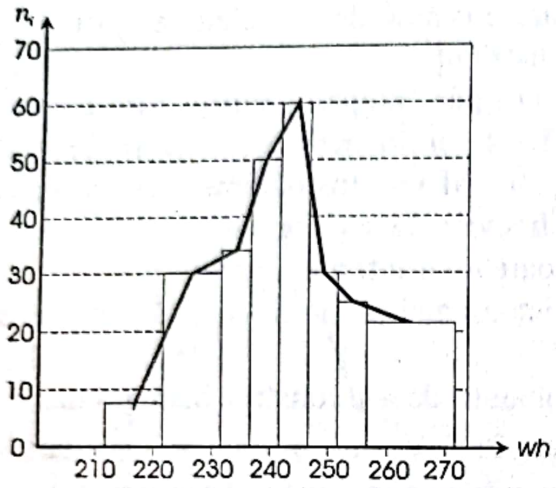


Figura 5.6. Distribuţia iniţială

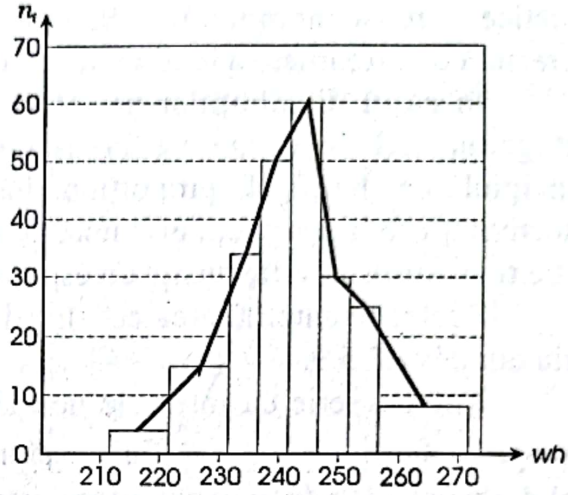


Figura 5.7. Distribuţia transformată

În histograma din figura 5.6 nu se respectă principiul proporţionalităţii dintre suprafaţa histogramei şi volumul colectivităţii. Prin reprezentarea frecvenţelor reale, suprafeţele dreptunghiurilor sunt majorate de un număr de ori , egal cu raportul dintre mărimea fiecărui interval şi mărimea celui mai mic interval.

Prin transformarea frecvenţelor (coloanele 3-5 ale tabelului 5.9) şi prin construirea histogramei pe baza lor, se respectă principiul mai sus amintit (fig.5.7).

Pentru construirea poligonului frecvenţelor şi a curbei frecvenţelor în cazul distribuţiilor pe intervale neegale, se procedează la fel ca şi în cazul distribuţiilor pe intervale egale, cu singura deosebire că, de data aceasta, se iau în considerare frecvenţele transformate în locul celor reale.

Curba frecvenţelor cumulate sau *ogiva* reprezintă una dintre cele mai clare şi utile reprezentări grafice ale seriilor de repartiţie. Frecvenţa cumulată este analogă cu funcţia empirică de repartiţie. Pentru construcţia ogivei reluăm exemplul din tabelul 5.9.

Scanned with OKEN Scanner

Tabelul 5.10. Distribuţia muncitorilor după productivitatea zilnică a muncii

Numărul de piese lucrate într-o zi	Numărul muncitorilor	N_i	N_i'
210-215	3	3	260
215-220	5	8	257
220-225	10	18	252
225-230	20	38	242
230-235	35	73	222
235-240	50	123	187
240-245	60	183	137
245-250	30	213	77
250-255	25	238	47
255-260	10	248	22
260-265	8	256	12
265-270	4	260	4
Total	260	-	-

Cele două frecvenţe cumulate N_i şi N_i' din tabelul 5.10, se reprezintă grafic prin ogivele din figura 5.8 şi figura 5.9.

Între histograma unei serii şi curbele cumulative există o strânsă legătură, după cum se observă din figura 5.10. Atât histograma, cât şi curbele cumulative au aceeaşi abscisă. Ca şi în cazul histogramei, suprafaţa fiecărui dreptunghi, din cadrul ogivelor, este proporţională cu mărimea frecvenţei intervalului respectiv. Datorită caracterului ogivei, baza fiecărui dreptunghi are ca ordonată frecvenţa cumulată a tuturor grupelor anterioare (ţinând seama de sensul cumulării).

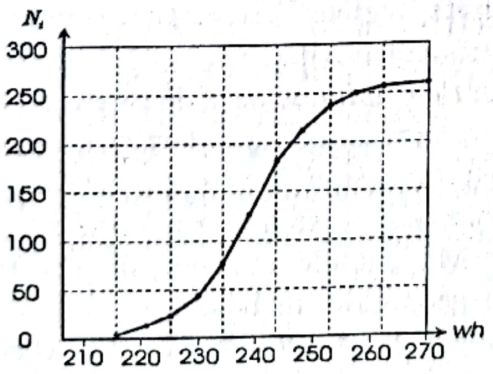


Figura 5.8. Poligonul frecvenţelor cumulate

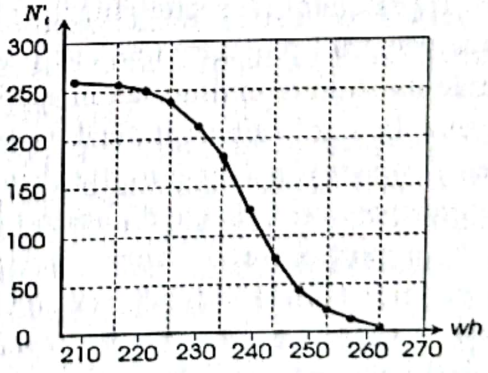


Figura 5.9. Poligonul frecvenţelor retrocumulate

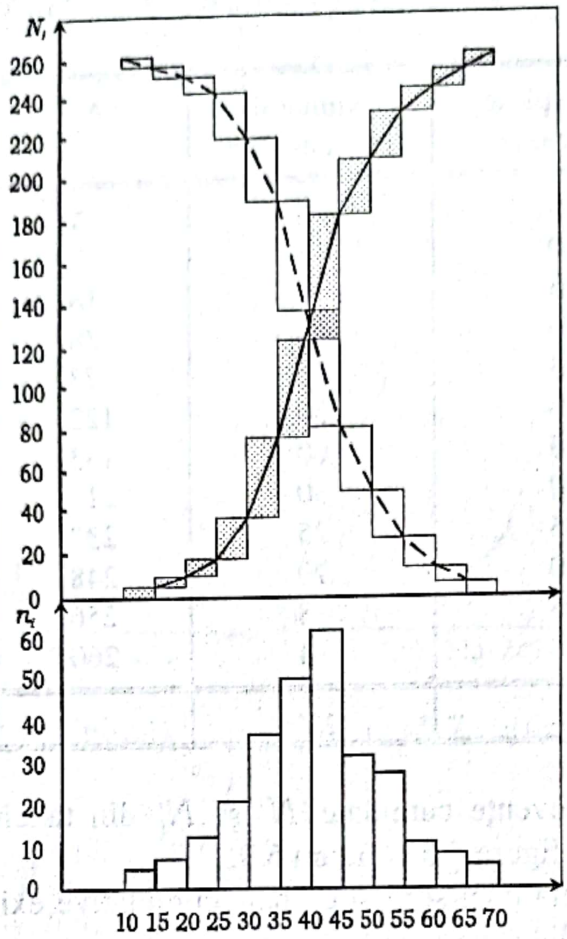


Figura 5.10. Histograma şi curbele frecvenţelor cumulate

Curba cumulativă creează impresia unei mai pronunţate continuităţi în dezvoltarea fenomenelor reprezentate, decât histograma sau poligonul frecvenţelor.

Diagramele de structură reprezintă forme speciale de grafice construite cu ajutorul suprafeţelor (dreptunghiuri, pătrate, cercuri, semicercuri etc.) şi se utilizează pentru reprezentarea structurii colectivităţilor şi a modificărilor structurale intervenite în timp sau în spaţiu. *Principiul* de bază de care trebuie să se ţină seama la construirea graficelor de structură *este proporţionalitatea între mărimea fenomenului cercetat şi mărimea suprafeţei figurii geometrice folosită pentru reprezentare*. Grupele din cadrul colectivităţii se reprezintă prin porţiuni de suprafaţă proporţionale cu volumul lor. Modalităţile concrete de construire a acestor diagrame diferă în funcţie de figura geometrică de bază.

În cazul folosirii *diagramelor prin dreptunghiuri*, suprafaţa acestora trebuie să fie direct proporţională cu volumul colectivităţilor cercetate, iar grupele din cadrul

Scanned with OKEN Scanner

colectivităţii se reprezintă prin porţiuni proporţionale din suprafaţa dreptunghiurilor. Raportul dintre suprafeţele acestor porţiuni şi suprafaţa totală a dreptunghiului trebuie să fie egal cu raportul dintre volumul fiecărei grupe şi volumul total al colectivităţii.

Tabelul 5.11. Populaţia lumii în anii 1950 şi 1990

	1950			1990		
	Milioane locuitori	%	Grade	Milioane locuitori	%	Grade
Populaţia mondială	2477	100	360	5295	100	360
- Regiuni dezvoltate	849	34	122	1211	23	83
- *populaţia urbană*	439	18	65	881	17	61
- *populaţia rurală*	410	16	57	330	6	22
- Regiuni slab dezvoltate	1628	66	238	4084	77	277
- *populaţia urbană*	265	11	40	1401	26	93
- *populaţia rurală*	1363	55	198	2683	51	184

Sursa: Prelucrat după: *"Trends in developing countries", World Bank*, 1973, şi *United Nations: "World Urbanization Prospects 1990"*, New York, p.

Pentru ilustrarea metodei de construire a *dreptunghiului de structură*, considerăm datele din tabelul 5.11, referitoare la structura populaţiei lumii după nivelul de dezvoltare şi după gradul de urbanizare în anii 1950 şi 1990. Reprezentarea grafică va trebui să evidenţieze atât creşterea populaţiei totale în decursul perioadei, cât şi modificările de structură care au avut loc în cele patru decenii (fig.5.11).

Considerăm două dreptunghiuri cu aceeaşi bază mică, iar pentru a asigura proporţionalitatea cu indicatorii pe care-i reprezintă, este suficient ca baza mare să fie proporţională cu indicatorii respectivi. Astfel, raportul dintre cele două baze mari va trebui să fie egal cu raportul dintre volumele celor două colectivităţi:

inv. 14970

$$5295:2477 \approx 2$$

Scanned with OKEN Scanner

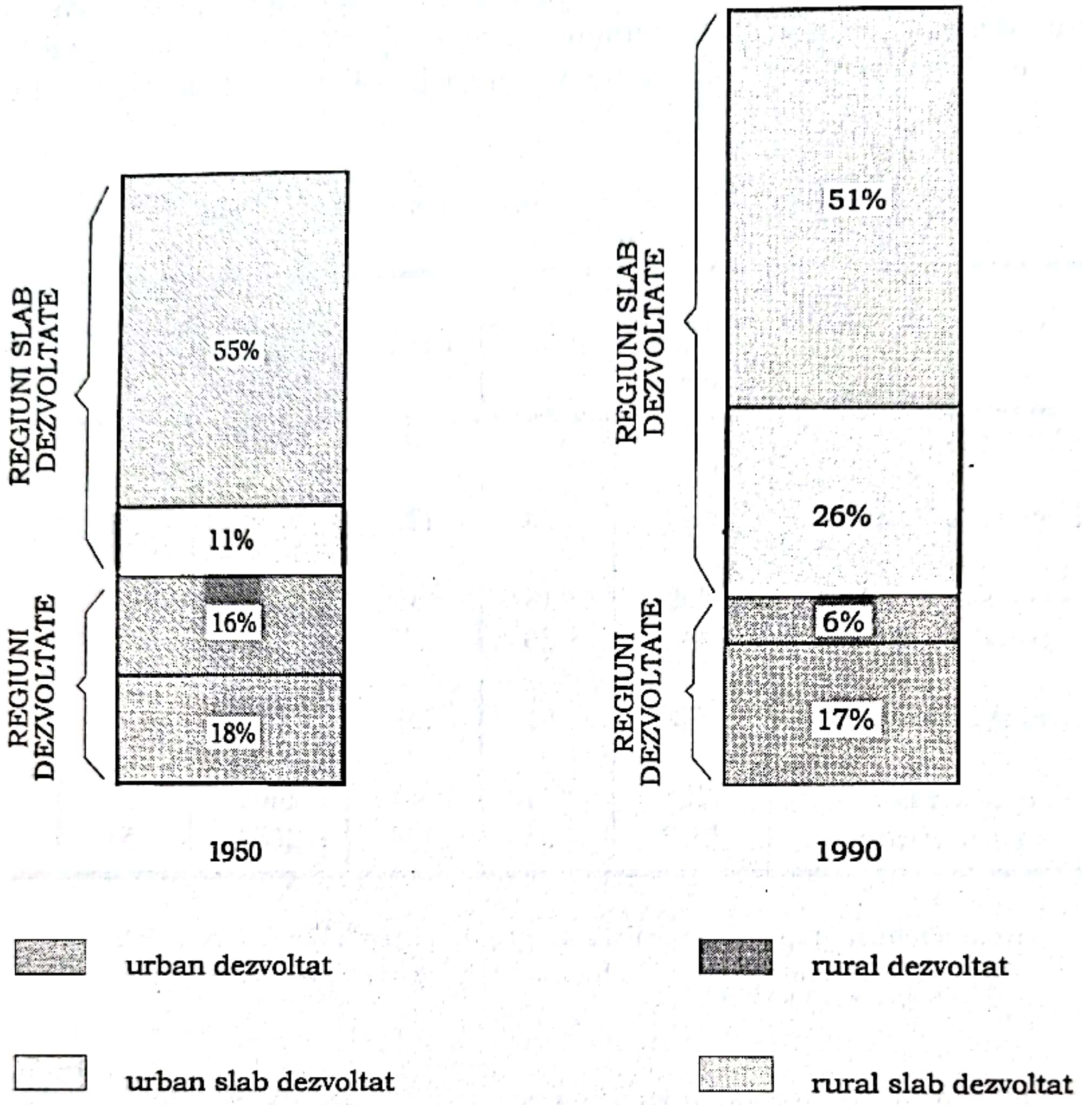


Figura 5.11. Populaţia lumii. Dreptunghiul de structură

Diagramele de structură prin pătrate se construiesc pe baza respectării aceluiaşi principiu al proporţionalităţii între suprafaţa pătratului şi volumul colectivităţii prezentate. Pentru aceasta se consideră un pătrat care reprezintă colectivitatea (100%) şi apoi acesta se împarte în 100 de pătrăţele, fiecare corespunzând unui procent. În continuare, pentru fiecare grupă a colectivităţii, se separă un număr de pătrăţele egal cu ponderea procentuală a grupei în totalul colectivităţii.

Scanned with OKEN Scanner

Revenind la exemplul din tabelul 5.11, se determină laturile celor două pătrate prin extragerea rădăcinii pătrate din volumul colectivităţilor în anii 1950 şi 1990:

$$S_{1950} = l^2_{1950} = 2477 \qquad \text{de unde} \qquad l_{1950} = \sqrt{2477} \cong 50$$
$$S_{1990} = l^2_{1990} = 5295 \qquad\qquad l_{1990} = \sqrt{5295} \cong 73$$

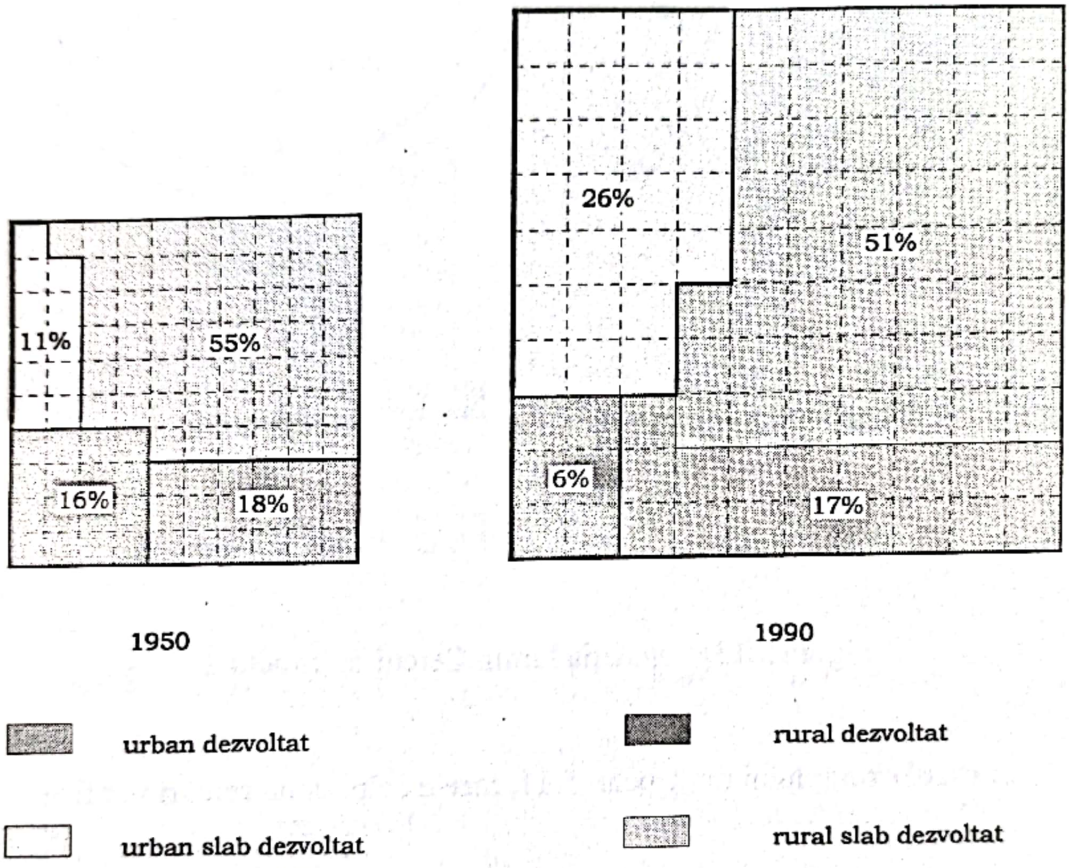


Figura 5.12. Populaţia lumii. Pătratul de structură

12Pentru lungimea laturii pătratului, rezultată din acest calcul, se alege unitatea de măsură convenabilă, în cazul de faţă (mm), figura 5.12.

Diagrama de structură prin cerc constă în reprezentarea structurii colectivităţilor prin sectoare de cerc. Suprafaţa cercului întreg va fi proporţională cu volumul colectivităţii de reprezentat:

$$S = \pi R^2 \qquad \text{de unde} \qquad R = \sqrt{\frac{S}{\pi}}$$

Scanned with OKEN Scanner

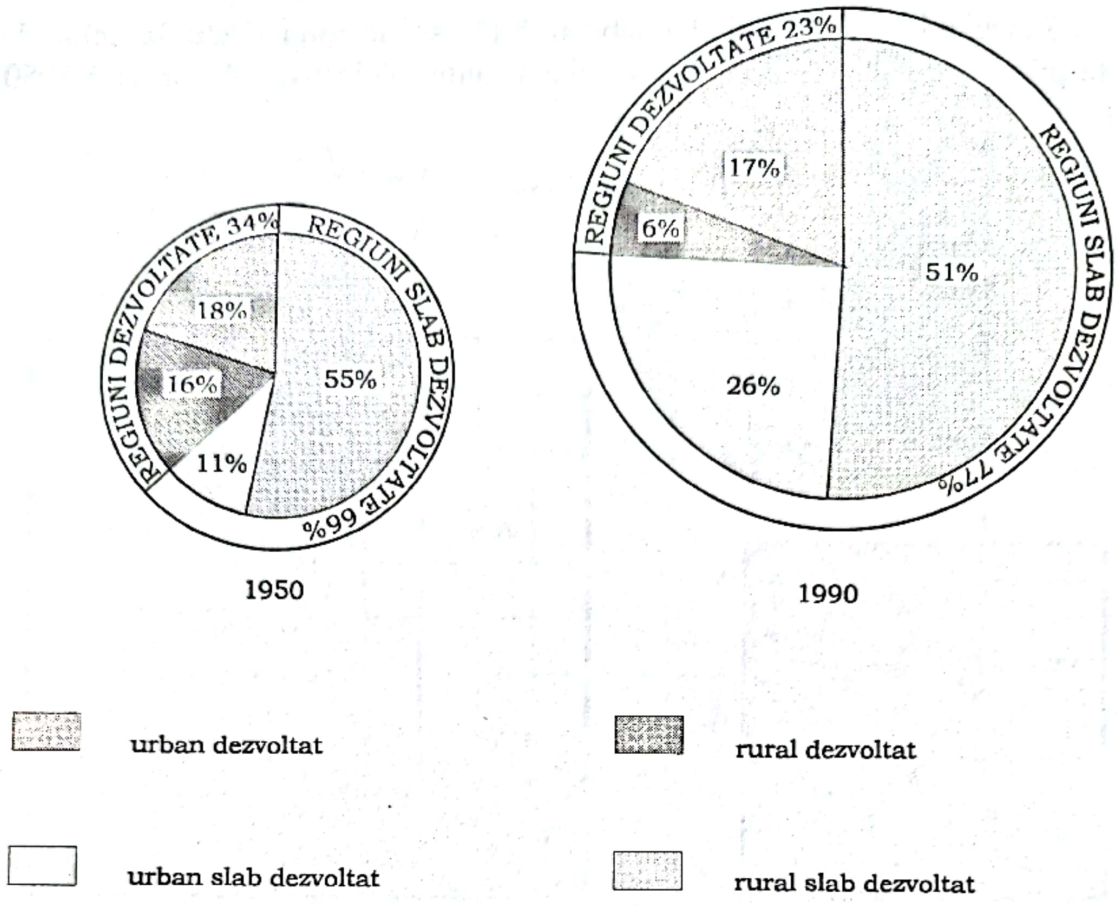


Figura 5.13. Populaţia lumii. Cercul de structură

În cazul exemplului din tabelul 5.11, razele celor două cercuri vor fi egale cu:

$$S_{1950} = 2477 \qquad R_{1950} = \sqrt{\frac{2477}{3{,}14}} \cong 28$$

iar

$$S_{1990} = 5295 \qquad R_{1990} = \sqrt{\frac{5295}{3{,}14}} \cong 41$$

Ca unitate de lungime considerăm mm. Pentru a împărţi fiecare cerc în sectoare proporţionale cu volumul grupelor din cadrul colectivităţilor,este suficient să se determine mărimea unghiului la centru, echivalând volumul colectivităţii 100% cu 360° (vezi tabelul 5.11, coloanele 4 şi 7). Mărimea unghiului pentru fiecare sector s-a determinat pe baza regulei de trei simplă.

Scanned with OKEN Scanner

Datorită faptului că populaţia mondială este grupată în două trepte, vom construi câte un cerc de arie proporţională cu volumul colectivităţii, pe care-l vom împărţi la început în două sectoare proporţionale cu populaţia din zonele dezvoltate şi zonele slab dezvoltate. Apoi vom construi un alt cerc, concentric cu primul, şi cele două sectoare determinate iniţial vor fi împărţite în câte două părţi, corespunzătoare mediului rural şi celui urban.

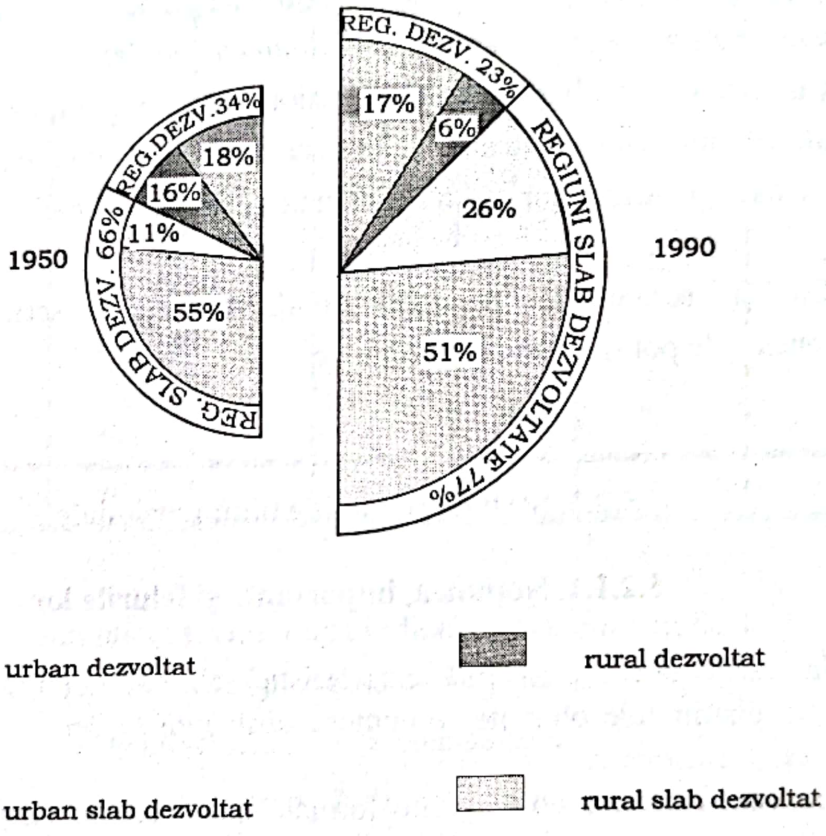


Figura 5.14. Populaţia lumii. Semicercul de structură

Un alt tip de grafic de structură este şi *semicercul de structură*. Principiul de construcţie al acestei diagrame este analog cu cel utilizat pentru construirea cercului de structură cu deosebirea că, în acest caz, dimensiunile unghiurilor pentru fiecare sector de cerc se iau pe jumătate, întrucât volumul colectivităţii se echivalează cu 180° (fig.5.14).

Scanned with OKEN Scanner

5.2. Serii de repartiţie multidimensionale

Pe lângă distribuţiile unidimensionale, în statistică se utilizează numeroase distribuţii după două, trei sau mai multe caracteristici. Deoarece frecvenţele în acest caz se referă la mai multe caracteristici, distribuţiile respective se numesc *distribuţii multidimensionale* sau *serii de repartiţie multidimensionale.*

Cu toate că distribuţiile multidimensionale înglobează, ca un caz particular, distribuţiile unidimensionale, în studiul lor apar şi anumite probleme specifice care nu au sens în cazul distribuţiilor unidimensionale, cum ar fi analiza legăturilor dintre fenomene.

Ca şi seriile de repartiţie unidimensionale, seriile de repartiţie multidimensionale pot fi teoretice şi empirice.

5.2.1. Serii de repartiţie bidimensionale

5.2.1.1. Noţiunea, importanţa şi felurile lor

În cazul studierii a două caracteristici ale unităţilor statistice ale unei colectivităţi, distribuţiile obţinute se numesc *distribuţii bidimensionale* sau *serii de repartiţie bidimensionale.*

Să considerăm o colectivitate formată din n unităţi statistice, care poartă simultan două caracteristici X şi Y şi să notăm cu $x_1, x_2, x_3, \ldots, x_i, \ldots, x_k$ cele k variante ale caracteristicii X şi cu $y_1, y_2, y_3, \ldots, y_i, \ldots, y_p$ cele p variante ale caracteristicii Y. Fie n_{ij} numărul de unităţi ale colectivităţii care poartă în acelaşi timp varianta x_i a caracteristicii X şi varianta y_j a caracteristicii Y, deci n_{ij} este frecvenţa absolută, iar f_{ij} frecvenţa relativă de la intersecţia rândului i cu coloana j.

Distribuţiile bidimensionale se prezintă sub forma unui tabel cu dublă intrare (tabelul 5.12), în care pe linii se înscriu variantele variabilei X, iar pe coloane cele ale variabilei Y, obţinându-se astfel un tabel cu k linii şi p coloane. Pentru caracteristicile X şi Y, suma frecvenţelor absolute n_{ij} este egală cu volumul colectivităţii:

Scanned with OKEN Scanner

$$\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} = n_{..}$$

Tabelul 5.12 Schema generală a unui tabel cu dublă intrare

X \ Y	y_1	y_2	...	y_j	...	y_p	Total
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kp}	$n_{k.}$
Total	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.p}$	$n_{..}$

Notăm printr-un punct totalul după indicele i sau după indicele j, astfel:

$n_{i.}$ - reprezintă totalul frecvenţelor absolute n_{ij} însumate după j;

$n_{.j}$ - totalul frecvenţelor absolute n_{ij} însumate după i;

$n_{..}$ totalul frecvenţelor absolute n_{ij} însumate concomitent după i şi j.

Din tabel rezultă că, deoarece variantele x_i şi y_i sunt incompatibile şi exhaustive, între frecvenţele absolute au loc relaţiile:

$$n_{i.} = \sum_{j=1}^{p} n_{ij} \quad \text{şi} \quad n_{.j} = \sum_{i=1}^{k} n_{ij} \qquad i = \overline{1,k}$$

$$n_{..} = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{k} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{p} n_{.j} = n \qquad j = \overline{1,p}$$

Această relaţie exprimă proprietatea de *comutativitate a sumei.*

În mod analog, suma tuturor frecvenţelor relative f_{ij} ale tuturor perechilor de valori corelate este egală cu unitatea:

Scanned with OKEN Scanner

$$f_{i\cdot} = \sum_{i=1}^{k} f_{ij} \quad \text{şi} \quad f_{\cdot j} = \sum_{j=1}^{p} f_{ij}$$

$$f_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{p} f_{ij} = \sum_{i=1}^{k} f_{i\cdot} = \sum_{j=1}^{p} f_{\cdot j} = 1 \qquad i = \overline{1,k} \quad j = \overline{1,p}$$

Dacă fiecare pereche de valori corelate (x_i, y_i) se întâlneşte o singură dată, atunci distribuţiile bidimensionale se prezintă sub forma a două şiruri paralele de valori, câte unul pentru fiecare variabilă.

Frecvenţele de pe ultima coloană şi de pe ultimul rând ale tabelului cu dublă intrare poartă denumirea de *frecvenţe marginale.*

O anumită variabilă împreună cu frecvenţele marginale corespunzătoare formează o *distribuţie marginală.* În cazul distribuţiilor bidimensionale există două distribuţii marginale: $(x_i, n_{i\cdot})$ sau $(x_i, f_{i\cdot})$ şi $(y_j, n_{\cdot j})$ sau $(y_j, f_{\cdot j})$.

Frecvenţele relative marginale se determină astfel:

$$f_{i\cdot} = \frac{n_{i\cdot}}{n_{\cdot\cdot}} \quad \text{şi} \quad f_{\cdot j} = \frac{n_{\cdot j}}{n_{\cdot\cdot}} \qquad \begin{matrix} i = \overline{1,k} \\ j = \overline{1,p} \end{matrix}$$

Distribuţiile marginale sunt distribuţii unidimensionale şi se bucură de aceleaşi proprietăţi ca şi acestea.

Pe lângă distribuţiile marginale, în cadrul unui tabel cu dublă intrare, se formează şi *distribuţii condiţionate* ale celor două variabile. Să considerăm cele $n_{\cdot j}$ unităţi statistice care poartă varianta y_j a caracteristicii Y (tabelul 5.12). Coloana j din tabelul cu dublă intrare se referă la o subcolectivitate din punct de vedere al caracteristicii X, formată din $n_{\cdot j}$ unităţi statistice. Se poate deduce că frecvenţa absolută a variantei x_i condiţionată de y_j, notată cu n_i^j este dată de relaţia:

$$n_i^j = n_{ij} \qquad i = \overline{1,k} \quad j = \overline{1,p}$$

iar frecvenţele relative condiţionate vor fi:

$$f_i^j = \frac{n_{ij}}{n_{\cdot j}}$$

Scanned with OKEN Scanner

Tabelul 5.13. Distribuţia caracteristicii X condiţionată de y_j

Caracteristica X	Frecvenţele absolute condiţionate n_i^j	Frecvenţele relative condiţionate f_i^j
x_1	n_1^j	f_1^j
x_2	n_2^j	f_2^j
$\vdots$	$\vdots$	$\vdots$
x_i	n_i^j	f_i^j
$\vdots$	$\vdots$	$\vdots$
x_k	n_k^j	f_k^j
Total	$n_{\cdot j}$	$f_{\cdot j}$

Frecvenţele absolute n_i^j şi frecvenţele relative f_i^j definesc distribuţia condiţionată după caracteristica X a unităţilor care poartă în acelaşi timp varianta y_j a caracteristicii Y. Evident, distribuţiile condiţionate, la fel ca şi distribuţiile marginale, sunt tot distribuţii unidimensionale. Din tabelul 5.12 se observă că sunt p distribuţii condiţionate ale caracteristicii X în raport cu Y (câte una pentru fiecare variantă y_j a caracteristicii Y) şi, în mod analog, vor fi k distribuţii condiţionate ale caracteristicii Y în raport cu X (câte una pentru fiecare variantă x_i caracteristicii X). Analog, frecvenţa relativă a lui y_j condiţionată de x_i, va fi:

$$f_i^j = \frac{n_{ij}}{n_{i\cdot}} \qquad i = \overline{1,k} \quad j = \overline{1,p}$$

Tot din definiţiile frecvenţelor condiţionate rezultă relaţiile:

$$\frac{n_{ij}}{n_{\cdot\cdot}} = \frac{n_{i\cdot} \cdot n_{ij}}{n_{\cdot\cdot} \cdot n_{i\cdot}} = \frac{n_{\cdot j} \cdot n_{ij}}{n_{\cdot\cdot} \cdot n_{\cdot j}} \quad \text{deci} \quad f_{ij} = f_{i\cdot} \cdot f_j^i = f_{\cdot j} \cdot f_i^j$$

Scanned with OKEN Scanner

Tabelul 5.14. Distribuţia variabilei Y condiţionată de x_i

Caracteristica Y	Frecvenţele absolute condiţionate n_j^i	Frecvenţele relative condiţionate f_j^i
y_1	n_1^i	f_1^i
y_2	n_2^i	f_2^i
⋮	⋮	⋮
y_j	n_j^i	f_j^i
⋮	⋮	⋮
y_p	n_p^i	f_p^i
Total	$n_{i\cdot}$	$f_{i\cdot}$

5.2.1.2. Reprezentarea grafică a seriilor de repartiţie bidimensionale

Distribuţia unităţilor statistice ale unei colectivităţi după două caracteristici se poate reprezenta grafic în spaţiul bidimensional sau în cel tridimensional, în timp ce distribuţiile unidimensionale se reprezintă în plan.

Pentru reprezentarea grafică a distribuţiilor bidimensionale se pot utiliza diferite figuri geometrice, cu scopul de a reda repartiţia în ansamblu prin puncte, suprafeţe, cercuri, paralelipipede dreptunghiulare etc.

Un grafic sugestiv, simplu de construit şi de aceea des utilizat în practica statistică este *diagrama prin puncte*, numită şi *diagrama* sub forma unui "*nor de puncte*". Pentru construirea acestui grafic, se ia un sistem de axe rectangulare pe care se înscriu cele două caracteristici X şi Y, iar fiecare pereche de valori $(x_i; y_i)$ se reprezintă printr-un punct. Dacă variaţia celor două caracteristici se prezintă sub formă de intervale, atunci în planul xoy se construieşte o reţea formată din perpendicularele ridicate pe cele două axe din punctele corespunzătoare limitelor intervalelor. Apoi în fiecare pătrat (sau dreptunghi) al reţelei se construieşte un număr de puncte egal cu frecvenţa corespunzătoare.

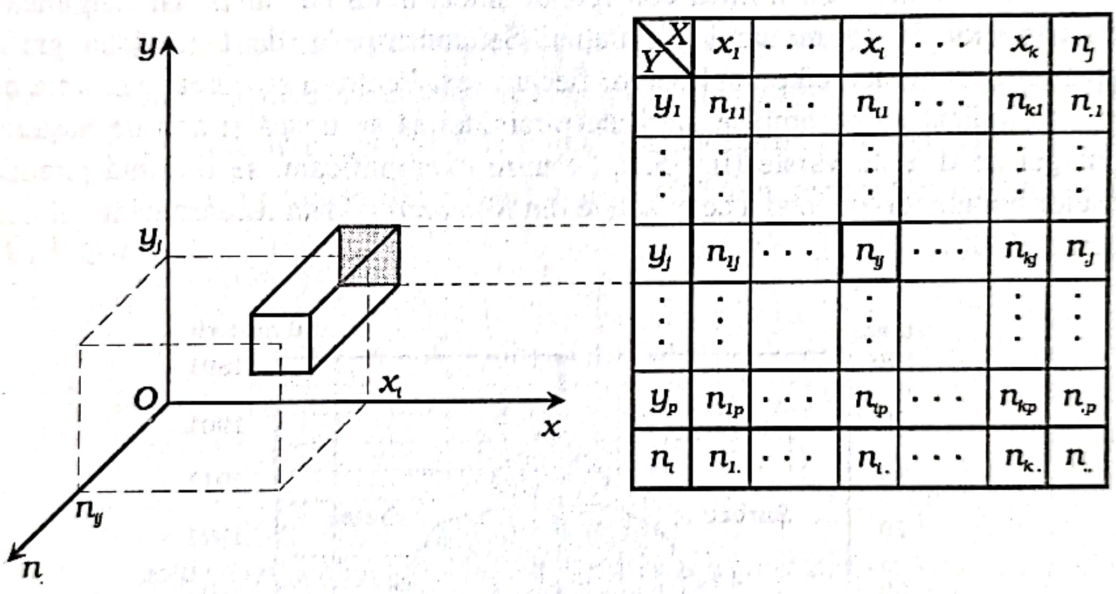
Concentrarea acestor puncte într-o anumită zonă din planul axelor de coordonate ia forma unui nor, de unde derivă şi denumirea de *nor de puncte*.

Cu ajutorul acestui grafic se poate aprecia existenţa unei legături dintre cele două variabile X şi Y, iar în funcţie de forma norului de puncte se apreciază forma şi intensitatea acesteia. Astfel, dacă punctele se dispersează fără nici o regulă în întreg

Scanned with OKEN Scanner

planul *xoy*, înseamnă că informaţiile empirice pe care le-am reprezentat grafic nu ne spun nimic cu privire la existenţa vreunei legături între cele două variabile. Dimpotrivă, concentrarea punctelor empirice într-o anumită zonă a graficului constituie un argument în favoarea ipotezei că între variabilele reprezentate poate exista o anumită legătură.

Un alt tip de grafic folosit pentru reprezentarea distribuţiilor bidimensionale îl constituie *diagrama prin paralelipipede dreptunghiulare*. Construcţia acestui grafic se prezintă schematic în figura 5.15. Se consideră un sistem de axe tridimensional corespunzător celor două caracteristici X şi Y şi frecvenţele perechilor de valori corelate n_{ij}. În planul *xoy* se trasează reţeaua ca şi în cazul anterior şi apoi, în fiecare pătrat (sau dreptunghi) din reţea, se construieşte câte un paralelipiped a cărui înălţime este proporţională cu frecvenţa considerată.



Y \ X	x_1	...	x_i	...	x_k	n_j
y_1	n_{11}	...	n_{i1}	...	n_{k1}	$n_{.1}$
⋮	⋮		⋮		⋮	⋮
y_j	n_{1j}	...	n_{ij}	...	n_{kj}	$n_{.j}$
⋮	⋮		⋮		⋮	⋮
y_p	n_{1p}	...	n_{ip}	...	n_{kp}	$n_{.p}$
n_i	$n_{1.}$	...	$n_{i.}$	...	$n_{k.}$	$n_{..}$

Figura 5.15. Construirea unei diagrame prin paralelipipede dreptunghiulare

Diagrama prin paralelipipede dreptunghiulare corespunde histogramei din cazul seriilor de distribuţie unidimensionale.

În mod analog cu poligonul de frecvenţă, din cazul distribuţiilor unidimensionale, repartiţia frecvenţelor după două caracteristici se poate reprezenta printr-o suprafaţă în spaţiul tridimensional şi poartă denumirea de *suprafaţă poliedrală*,

Scanned with OKEN Scanner

sau *poliedru al frecvenţelor*. Poliedrul frecvenţelor se obţine reprezentând frecvenţele perechilor de valori $(x_i; y_i)$ prin segmente proporţionale cu acestea şi ridicate perpendicular pe planul axelor de coordonate *xoy* din centrul fiecărui pătrat sau romb al reţelei graficului. Unind consecutiv extremităţile de sus ale perpendicularelor, se obţine poliedrul frecvenţelor.

Un alt mod de reprezentare grafică a distribuţiilor bidimensionale este stereograma, care constă în reprezentarea în plan a frecvenţelor perechilor de valori $(x_i; y_i)$ prin cercuri de arii proporţionale cu acestea. Centrul fiecărui cerc coincide cu punctul de intersecţie al diagonalelor din reţeaua graficului *xoy*.

Un exemplu clasic de reprezentare grafică a distribuţiilor bidimensionale îl constituie *piramida vârstelor*. Cu ajutorul ei se poate reprezenta distribuţia populaţiei pe vârste şi sexe la un anumit moment. Pentru a construi un astfel de grafic pe axa verticală se trece vârsta populaţiei de la 0 la 100 de ani, iar pe orizontală, numărul populaţiei. Numărul populaţiei de diferite vârste sau grupe de vârste se reprezintă printr-o bandă a cărei bază mică corespunde intervalului de vârstă, iar lungimea ei este proporţională cu numărul populaţiei. Se construiesc, de fapt, două grafice (histograme) alăturate, câte unul pentru fiecare sex. Pentru a se putea interpreta mai uşor este indicat ca pe ambele părţi ale piramidei să se treacă şi anii de naştere a populaţiei de diferite vârste (fig. 5.16). Pentru exemplificare, se prezintă piramida vârstelor populaţiei totale şi a celei active din *România* la data recensământului din 7 ianuarie 1992.

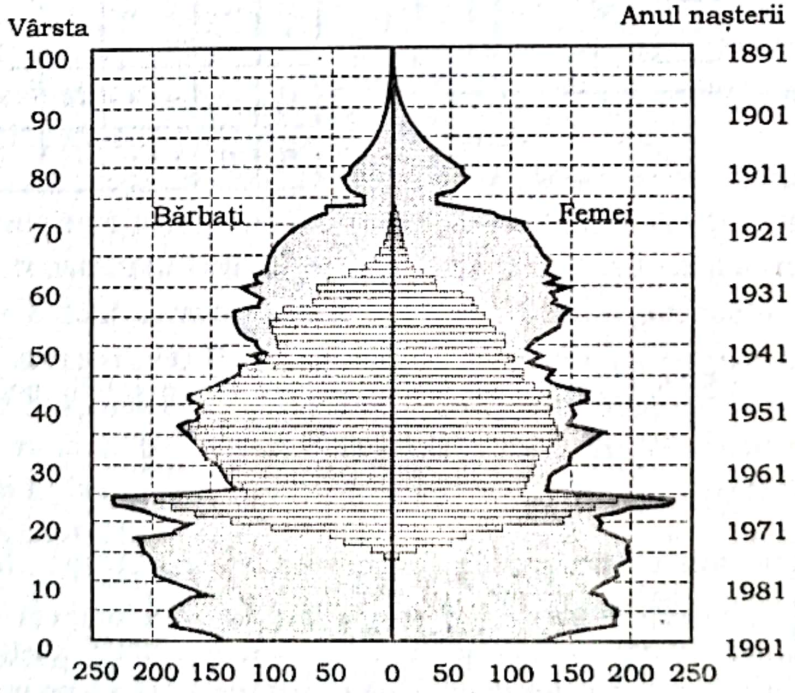


Figura 5.16. Populaţia activă şi cea inactivă din *România* pe vârste şi sexe la recensământul din 7 ianuarie 1992

Scanned with OKEN Scanner

Piramida arată sugestiv existenţa a şase *intrânduri* (îngustări) la ambele sexe, ca urmare a efectelor negative ale celor două războaie mondiale şi a evoluţiei contradictorii a natalităţii după anul 1956. Reducerea natalităţii în timpul celor două războaie mondiale: (1916-1919 şi 1939-1945) asociată cu creşterea însemnată a mortalităţii infantile se reflectă în cele două intrânduri ale piramidei din partea ei superioară şi mediană, iar scăderea natalităţii după anul 1956 determină cea de-a treia îngustare în partea mediană a ei, în timp ce tendinţa de scădere a natalităţii de după 1970 se reflectă în îngustarea tot mai pronunţată a bazei piramidei.

5.2.2. Repartiţii tridimensionale

Distribuţiile tridimensionale rezultă din considerarea a trei caracteristici cantitative sau calitative sau cantitative şi calitative pentru fiecare unitate statistică.

În cazul unui număr redus de unităţi statistice, distribuţiile tridimensionale se prezintă sub forma a trei şiruri de date paralele, câte unul pentru fiecare caracteristică X, Y şi Z:

$$X: \quad (x_1, x_2, \ldots, x_i, \ldots, x_n)$$
$$Y: \quad (y_1, y_2, \ldots, y_i, \ldots, y_n)$$
$$Z: \quad (z_1, z_2, \ldots, z_i, \ldots, z_n)$$

Aceste şiruri de date pot fi eventual ordonate crescător sau descrescător, după variantele uneia dintre cele trei variabile.

Ca şi în cazul distribuţiilor bidimensionale, când numărul unităţilor statistice n este suficient de mare, se recomandă sistematizarea datelor iniţiale sub forma unei distribuţii de frecvenţă a celor trei variabile pe variante sau pe grupe de variante ale caracteristicilor. Se obţine, astfel, un tabel statistic cu triplă intrare (tabelul 5.15) care cuprinde frecvenţele n_{ijk} ale ansamblului de variante (x_i, y_j, z_k), $(i = 1,2,\ldots,p; \quad j = 1,2,\ldots,q; \quad k = 1,2,\ldots,r)$ şi frecvenţele marginale $n_{.jk}$, $n_{i.k}$, $n_{.j.}$ şi $n_{.k.}$.

Scanned with OKEN Scanner

5.15. Schema generală a unui tabel cu triplă intrare

X	Z / Y	z_1	z_2	...	z_k	...	z_r	Total
x_1	y_1	n_{111}	n_{112}	...	n_{11k}	...	n_{11r}	$n_{11.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_j	n_{1j1}	n_{1j2}	...	n_{1jk}	...	n_{1jr}	$n_{1j.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_q	n_{1q1}	n_{1q2}	...	n_{1qk}	...	n_{1qr}	$n_{1q.}$
	Total	$n_{1.1}$	$n_{1.2}$	...	$n_{1.k}$	...	$n_{1.r}$	$n_{1..}$
⋮	⋮	⋮	⋮	...	⋮	...	⋮	⋮
x_i	y_1	n_{i11}	n_{i12}	...	n_{i1k}	...	n_{i1r}	$n_{i1.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_j	n_{ij1}	n_{ij2}	...	n_{ijk}	...	n_{ijr}	$n_{ij.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_q	n_{iq1}	n_{iq2}	...	n_{iqk}	...	n_{iqr}	$n_{iq.}$
	Total	$n_{i.1}$	$n_{i.2}$	...	$n_{i.k}$	...	$n_{i.r}$	$n_{i..}$
⋮	⋮	⋮	⋮	...	⋮	...	⋮	⋮
x_p	y_1	n_{p11}	n_{p12}	...	n_{p1k}	...	n_{p1r}	$n_{p1.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_j	n_{pj1}	n_{pj2}	...	n_{pjk}	...	n_{pjr}	$n_{pj.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_q	n_{pq1}	n_{pq2}	...	n_{pqk}	...	n_{pqr}	$n_{pq.}$
	Total	$n_{p.1}$	$n_{p.2}$	...	$n_{p.k}$	...	$n_{p.r}$	$n_{p..}$
Total	y_1	$n_{.11}$	$n_{.12}$	...	$n_{.1q}$	...	$n_{.1r}$	$n_{.1.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_j	$n_{.j1}$	$n_{.j2}$	...	$n_{.jk}$	...	$n_{.jr}$	$n_{.j.}$
	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
	y_q	$n_{.q1}$	$n_{.q2}$	...	$n_{.qk}$	...	$n_{.qr}$	$n_{.q.}$
TOTAL	Total	$n_{..1}$	$n_{..2}$	...	$n_{..k}$	...	$n_{..r}$	$n_{...}$

Scanned with OKEN Scanner

În cadrul tabelului 5.15 au loc relaţiile:

$$n_{...} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} n_{ijk} \qquad i=\overline{1,p} \qquad j=\overline{1,q} \qquad k=\overline{1,r}$$

$$n_{.11} = \sum_{i=1}^{p} n_{i11} \qquad n_{11.} = \sum_{k=1}^{r} n_{11k}$$

$$n_{.12} = \sum_{i=1}^{p} n_{i12} \qquad n_{21.} = \sum_{k=1}^{r} n_{21k}$$

$$\cdots\cdots\cdots\cdots \qquad \cdots\cdots\cdots\cdots$$

$$n_{.1r} = \sum_{i=1}^{p} n_{i1r} \qquad n_{p1.} = \sum_{k=1}^{r} n_{p1k}$$

$$n_{.1.} = \sum_{k=1}^{r} n_{.1k} \qquad n_{.1.} = \sum_{k=1}^{r} n_{i1k}$$

deci $$\sum_{k=1}^{r} n_{.1k} = \sum_{i=1}^{p} n_{i1.}$$

iar

$$n_{.21} = \sum_{i=1}^{p} n_{i21} \qquad si \qquad n_{12.} = \sum_{k=1}^{r} n_{12k}$$

$$n_{.22} = \sum_{i=1}^{p} n_{i22} \qquad n_{22.} = \sum_{k=1}^{r} n_{22k}$$

$$\cdots\cdots\cdots\cdots \qquad \cdots\cdots\cdots\cdots$$

$$n_{.2r} = \sum_{i=1}^{p} n_{i2r} \qquad n_{p2.} = \sum_{k=1}^{r} n_{p2k}$$

$$n_{.2.} = \sum_{k=1}^{r} n_{.2k} \qquad n_{.2.} = \sum_{i=1}^{p} n_{i2.}$$

Scanned with OKEN Scanner

deci $$\sum_{k=1}^{r} n_{.2k} = \sum_{i=1}^{p} n_{i2.}$$

$$n_{1..} = \sum_{k=1}^{r} n_{1.k} = \sum_{j=1}^{q}\sum_{k=1}^{r} n_{1jk} \qquad si \qquad n_{..1} = \sum_{j=1}^{q} n_{.j1} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij1}$$

$$n_{2..} = \sum_{k=1}^{r} n_{2.k} = \sum_{j=1}^{q}\sum_{k=1}^{r} n_{2jk} \qquad n_{..2} = \sum_{j=1}^{q} n_{.j2} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ij2}$$

....

$$n_{i..} = \sum_{k=1}^{r} n_{i.k} = \sum_{j=1}^{q}\sum_{k=1}^{r} n_{ijk} \qquad n_{..k} = \sum_{j=1}^{q} n_{.jk} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ijk}$$

....

$$n_{p..} = \sum_{k=1}^{r} n_{p.k} = \sum_{j=1}^{q}\sum_{k=1}^{r} n_{pjk} \qquad n_{..r} = \sum_{j=1}^{q} n_{.jr} = \sum_{i=1}^{p}\sum_{j=1}^{q} n_{ijr}$$

De unde rezultă:

$$n_{...} = \sum_{i=1}^{p} n_{i..} = \sum_{k=1}^{r} n_{..k} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} n_{ijk}$$

$$n_{...} = \sum_{j=1}^{q} n_{.j.} = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r} n_{ijk}$$

deci $$\sum_{i=1}^{p} n_{i..} = \sum_{j=1}^{q} n_{i..} = \sum_{k=1}^{r} n_{..k}$$

Din tabelul 5.15 se observă că, în cadrul distribuţiilor tridimensionale, se formează două tipuri de distribuţii marginale şi două tipuri de distribuţii condiţionate.

Scanned with OKEN Scanner

Considerăm, mai întâi, distribuţiile celor $n_{...} = n$ unităţi statistice în funcţie de câte o singură caracteristică X, Y sau Z.

Se formează astfel trei distribuţii marginale unidimensionale (tabelele 5.16, 5.17 şi 5.18).

Tabelul 5.16

X	$n_{i..}$
x_1	$n_{1..}$
x_2	$n_{2..}$
⋮	⋮
x_i	$n_{i..}$
⋮	⋮
x_p	$n_{p..}$
Total	$n_{...}$

Tabelul 5.17

Y	$n_{.j.}$
y_1	$n_{.1.}$
y_2	$n_{.2.}$
⋮	⋮
y_i	$n_{.j.}$
⋮	⋮
y_p	$n_{.p.}$
Total	$n_{...}$

Tabelul 5.18

Z	$n_{..r}$
z_1	$n_{..1}$
z_2	$n_{..2}$
⋮	⋮
z_i	$n_{..i}$
⋮	⋮
z_p	$n_{..p}$
Total	$n_{...}$

Se pot forma, de asemenea, trei distribuţii marginale bidimensionale, considerând variabilele perechi (X,Y), (X,Z) şi (Y,Z). Astfel distribuţia marginală a perechii de variabile Y şi Z se prezintă în tabelul 5.19.

Tabelul 5.19. Schema generală a distribuţiei marginale bidimensionale Y,Z

Y \ Z	z_1	...	z_k	...	z_r	Total
y_1	$n_{.11}$	...	$n_{.1k}$	...	$n_{.1r}$	$n_{.1.}$
y_2	$n_{.21}$	...	$n_{.2k}$	...	$n_{.2r}$	$n_{.2.}$
⋮	⋮	...	⋮	...	⋮	⋮
y_j	$n_{.j1}$	...	$n_{.jk}$	...	$n_{.jr}$	$n_{.j.}$
⋮	⋮	...	⋮	...	⋮	⋮
y_q	$n_{.q1}$	...	$n_{.qk}$	...	$n_{.qr}$	$n_{.q.}$
Total	$n_{..i}$	...	$n_{..k}$	...	$n_{..r}$	$n_{...}$

Scanned with OKEN Scanner

În mod analog, atenţia se poate fixa ca în tabelul 5.15, asupra distribuţiilor condiţionate de o singură valoare a variabilelor X, Y sau Z, care sunt distribuţii unidimensionale sau asupra distribuţiilor condiţionate de două valori ale variabilelor $(X,Y), (Y,Z), (X,Z)$, care sunt distribuţii bidimensionale.

Pentru exemplificare să considerăm distribuţia $(X, Y / Z_1)$ condiţionată de varianta z_1 care se prezintă în tabelul 5.20.

Tabelul 5.20. Distribuţia variabilelor x şi y condiţionate de z_1

Y \ X	x_1	x_2	...	x_i	...	x_p	Total
y_1	n_{111}	n_{211}	...	n_{i11}	...	n_{p11}	$n_{.11}$
y_2	n_{121}	n_{221}	...	n_{i21}	...	n_{p21}	$n_{.21}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_j	n_{1j1}	n_{2j1}	...	n_{ij1}	...	n_{pj1}	$n_{.j1}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
y_q	n_{1q1}	n_{2q1}	...	n_{iq1}	...	n_{pq1}	$n_{.q1}$
Total	$n_{1.1}$	$n_{2.1}$	...	$n_{i.1}$	...	$n_{p.1}$	$n_{..1}$

Distribuţia variabilei Z condiţionată de x_1 şi y_1, este tot o distribuţie unidimensională, de forma celei din tabelul 5.21.

Tabelul 5.21. Schema generală a unei distribuţii dublu condiţionate unidimensionale

Z	n_{11k}
z_1	n_{111}
z_2	n_{112}
⋮	⋮
z_k	n_{11k}
⋮	⋮
z_r	n_{11r}
Total	$n_{11.}$

Scanned with OKEN Scanner

5.3. Seriile dinamice

5.3.1. Noţiunea, importanţa şi felurile lor

Cunoaşterea ştiinţifică a evoluţiei fenomenelor în timp constituie o problemă complexă şi deosebit de importantă. Acest lucru se explică prin faptul că schimbările în timp care au loc în dezvoltarea colectivităţilor se referă la variaţia volumului şi structurii lor, la variaţia caracteristicilor proprii diferitelor unităţi ale lor şi la forma şi caracterul legăturii dintre aceste caracteristici. Toate aceste schimbări sunt, la rândul lor, interdependente.

Seriile dinamice, de timp sau *cronologice* se caracterizează prin faptul că termenii lor se diferenţiază în funcţie de o caracteristică de timp.

Ca orice serie statistică, şi cele dinamice se compun din două şiruri paralele de date. Primul se referă la variaţia caracteristicii t_i şi reprezintă diferite momente sau intervale de timp şi al doilea la nivelurile fenomenului studiat y_i, tabelul 5.22.

Tabelul 5.22. Prezentarea generală a unei serii dinamice

Caracteristica t_i	Nivelurile fenomenului y_i
t_0	y_0
t_1	y_1
t_2	y_2
⋮	⋮
t_{i-1}	y_{i-1}
t_i	y_i
t_{i+1}	y_{i+1}
⋮	⋮
t_n	y_n

O caracteristică esenţială a seriilor dinamice o constituie faptul că ele sunt *construite pentru colectivităţi omogene*. Aceasta asigură comparabilitatea indicatorilor de la o perioadă la alta sau de la un moment la altul. Seriile dinamice se pot construi pentru perioade de timp mai lungi sau mai scurte cu condiţia ca, în aceste

Scanned with OKEN Scanner

perioade, să nu fi avut loc modificări calitative importante în cadrul colectivităţii şi ele servesc pentru caracterizarea evoluţiei fenomenelor individuale sau colective pe diferite trepte sau etape din dezvoltarea lor.

O altă trăsătură caracteristică a seriilor dinamice constă în faptul *că fenomenele şi faptele sociale şi economice*, oglindite în indicatorii care formează seria, *sunt dependente între ele*. Nivelul unui fenomen, corespunzător unui anumit moment sau perioadă din dezvoltarea lui, depinde şi de nivelul aceluiaşi fenomen înregistrat în perioadele anterioare, ca şi de ambianţa în care a avut loc dezvoltarea.

De aceea, schemele teoretice ale calculului probabilităţilor - bazate, în general, pe independenţa fenomenelor - nu sunt aplicabile în analiza seriilor dinamice.

La rândul lor seriile dinamice, după natura colectivităţii la care se referă, se împart în două categorii şi anume: *serii de momente* şi *serii de intervale*.

Seriile de momente caracterizează variaţia în timp a colectivităţilor de stări, iar termenii lor reprezintă volumul acestor colectivităţi înregistrat în diferite momente succesive ale timpului. De exemplu: volumul stocurilor de produse existent în întreprinderi la începutul fiecărei luni, numărul populaţiei obţinut cu ocazia diferitelor recensăminte sau la data de 1 ianuarie a fiecărui an, numărul animalelor domestice, obţinut pe baza recensămintelor etc.

Să considerăm, pentru exemplificare, o serie care se referă la evoluţia populaţiei globului în era noastră (tabelul 5.23).

Tabelul 5.23. Evoluţia populaţiei globului în era noastră

Anii	Milioane locuitori	Anii	Milioane locuitori
1	100	1800	1265
1000	275	1900	2000
1100	306	1950	2486
1200	348	1960	2982
1300	384	1970	3632
1400	373	1980	4450
1500	446	1987	5000
1600	486	1990	5295
1650	500	1995	5742
1750	791	*2000*	*6228*[x)]

[x)] prognoză

Sursa: *ONU: "The Determinants and Consequences of Population Trends", vol.I*, New York, 1973, p.10.
ONU: World Population Prospects, The 1994 Revision, New York, 1995, p.216-217.

Scanned with OKEN Scanner

Din această serie dinamică se poate constata că populaţia globului a înregistrat în era noastră un proces de creştere continuă, mai lent la început, apoi din ce în ce mai accelerat până în prezent.

Deşi nu se precizează în tabel, se subînţelege că fiecare efectiv al populaţiei se referă la o anumită dată din fiecare an luat în calcul.

Seriile de momente se caracterizează prin faptul că *indicatorii lor succesivi nu se pot însuma*, o asemenea operaţie nu ar avea un conţinut real. Astfel, ar fi lipsită de conţinut şi în discordanţă cu realitatea însumarea numărului populaţiei lumii la diferite momente, deoarece o parte din populaţia care a fost luată în calcul, de exemplu, în anul 1995, exista şi în momentele precedente 1900, 1950, 1960, 1970, 1980 şi 1990.

O altă particularitate a seriilor dinamice de momente este şi aceea că ele se întocmesc pentru fenomenele a căror *înregistrare nu este posibil să se facă în mod curent-continuu*. Momentele unei astfel de serii pot fi egal sau neegal distanţate între ele şi această particularitate prezintă interes pentru alegerea celor mai potrivite metode şi tehnici de calcul şi analiză a seriilor de momente.

Seriile dinamice de intervale se referă la fenomenele care se înregistrează în mod curent şi cu caracter de continuitate pe măsura producerii lor, totalizându-se pentru anumite perioade de timp (zi, săptămână, lună, trimestru, semestru, an, cinci ani, zece ani etc.). De exemplu, producţia industrială, producţia agricolă, volumul vânzărilor de mărfuri, numărul născuţilor-vii, numărul căsătoriilor, numărul divorţurilor, numărul deceselor etc., se pot totaliza pe perioade de timp calendaristice. Pentru exemplificare se consideră seria din tabelul 5.24.

Tabelul 5.24. Locuinţele date în folosinţă în *România* în anii 1951 - 1995

Anii	Numărul locuinţelor
1951-1960	1 293 710
1961-1970	1 553 292
1971-1980	1 592 540
1981-1990	1 227 522
1991	27 958
1992	27 538
1993	30 071
1994	36 743
1995	35 822
1991-1995	158 132
1951-1995	5 825 196

Sursa: *Anuarele statistice ale României 1995*, pp.210-211 şi *1996*, pp.202-203

Scanned with OKEN Scanner

Deoarece seriile de intervale se obţin prin înregistrarea curentă-continuă a fenomenelor, indicatorii statistici care caracterizează nivelul dezvoltării lor pe o anumită perioadă de timp, se pot cumula din aproape în aproape, obţinându-se în felul acesta indicatori corespunzători unor intervale de timp mai mari. În cazul seriei din tabelul 5.24, prin cumularea numărului de locuinţe din perioadele 1951-1960, 1961-1970, 1971-1980, şi 1981-1990, s-a obţinut numărul locuinţelor date în folosinţă pe cincinale, pe decenii şi pe întreaga perioadă 1951-1995.

Mărimea intervalelor seriilor dinamice se stabileşte în funcţie de natura colectivităţilor şi de necesităţile practice ale analizei statistice.

Din punct de vedere al naturii indicatorilor, care reprezintă variantele caracteristicii cercetate, seriile dinamice pot fi de mai multe feluri: serii dinamice construite din mărimi absolute, serii dinamice construite din mărimi relative şi serii dinamice construite din mărimi medii.

Seriile dinamice construite din mărimi absolute caracterizează variaţia nivelului absolut al fenomenelor la anumite date sau pe anumite perioade de timp, mai mult sau mai puţin îndelungate, ele pot fi *de momente* sau *de intervale*.

Tabelul 5.25. Populaţia *României* în anii 1891-1996

Data	Numărul populaţiei
12 februarie	10 000 000
31 decembrie	11 168 000
31 decembrie	12 898 000 (recensământ)
31 decembrie	12 419 000
29 decembrie	14 280 129 (recensământ)
6 aprilie	16 126 063 (recensământ)
25 ianuarie	15 872 624 (recensământ)
21 februarie	17 489 450 (recensământ)
15 martie	19 103 163 (recensământ)
20 iunie	20 000 000
5 ianuarie	21 559 910 (recensământ)
1 iulie	22 201 387
3 decembrie	23 000 000
1 iulie	23 206 720
1 iulie	23 185 084
7 ianuarie	22 810 035 (recensământ)
1 iulie	22 755 260
1 iulie	22 730 622
1 ianuarie	22 656 145

Sursa: *Recensământul populaţiei şi locuinţelor din 15 martie 1966, vol.I,Partea I-a*, DCS, 1969, p.XVII; *Anuarul statistic al României 1995*, CNS, p.82

Scanned with OKEN Scanner

Să considerăm, pentru exemplificare, seria de momente reprezentând evoluţia numărului populaţiei *României* în perioada 1891-1996 (tabelul 5.25).

Din tabelul 5.25 se observă că numărul populaţiei *României* din anul 1891 s-a dublat în anul 1969, adică în decurs de 78 de ani. De asemenea, se remarcă ravagiile făcute de cele două războaie mondiale care au determinat câte un regres în cadrul tendinţei generale de creştere a populaţiei ţării până în anul 1991.

Tabelul 5.26. Dinamica produsului social şi a venitului naţional în perioada 1950-1990

Anii	Produsul social %	Venitul naţional %
1950	100	100
1960	263	268
1970	629	599
1980	15 ori	15 ori
1985	18 ori	17 ori
1986	19 ori	18 ori
1987	19 ori	18 ori
1988	19 ori	17 ori
1989	19 ori	16 ori
1990	16 ori	14 ori

Sursa: *Anuarul statistic al României 1991*, CNS, p.248-249.

Tabelul 5.27. Numărul mediu al salariaţilor din *România* în perioada 1950-1995

Anii	Mii salariaţi
1950	2123,0
1960	3249,2
1970	5108,7
1980	7378,0
1985	7700,3
1990	8156,0
1991	7574,0
1992	6888,0
1993	6672,0
1994	6438,0
1995	6160,0

Sursa: *Anuarul statistic al României 1995*, CNS, p. 152-153
Anuarul statistic al României 1996, CNS, p. 148-149

Scanned with OKEN Scanner

Seriile dinamice pot fi construite şi din *mărimi relative*, în raport cu o anumită bază fixă sau variabilă a fenomenelor în diferite momente sau perioade de timp. Prin urmare şi aceste serii pot fi *de intervale* sau *de momente*. De exemplu: dinamica produsului intern brut, produsului social, dinamica venitului naţional, dinamica productivităţii muncii pe un muncitor din industrie, dinamica populaţiei ocupate, dinamica numărului mediu scriptic al salariaţilor pentru diferite perioade de timp şi în funcţie de nivelul unui anumit indicator dintr-o anumită perioadă sau moment considerat ca bază etc.

Prezentăm, pentru concretizare, în paralel două serii dinamice ai căror termeni se exprimă în mărimi relative, şi anume dinamica produsului social şi a venitului naţional creat în *România* în perioada 1950-1990, în care nivelul anului 1950 este considerat ca bază (tabelul 5.26).

Seriile dinamice construite din *mărimi medii* caracterizează dezvoltarea medie a fenomenelor pe diferite trepte ale evoluţiei lor, de exemplu, evoluţia numărului mediu scriptic al salariaţilor din economia ţării noastre în perioada 1950-1995 (tabelul 5.27).

La elaborarea oricărei serii dinamice trebuie să se respecte anumite *condiţii esenţiale* care ar putea fi sistematizate astfel:

- *compararea* nivelului de dezvoltare a fenomenelor să se efectueze numai pentru perioade de timp egale;
- stabilirea *mărimii intervalului* de timp, în cazul seriilor de intervale, trebuie să se facă ţinând seama de natura specifică de dezvoltare a fenomenului respectiv, de schimbările calitative care se produc în cadrul colectivităţilor ca şi de necesităţile practice;
- să se asigure *comparabilitatea* indicatorilor seriilor dinamice din punct de vedere metodologic, teritorial, al preţurilor etc;
- orice serie dinamică trebuie să fie construită numai din *indicatori de acelaşi fel* (mărimi absolute, mărimi medii sau mărimi relative).

5.3.2. Reprezentarea grafică a seriilor dinamice

În practica statistică, în afară de forma obişnuită de prezentare a seriilor dinamice prin tabele se foloseşte, pe scară largă, reprezentarea lor grafică pentru a reda vizual aspectul dinamic, evolutiv al fenomenelor cercetate. Cu ajutorul

Scanned with OKEN Scanner

reprezentării grafice apare mai clar evoluţia fenomenelor în timp, poate fi efectuată compararea traiectoriilor mai multor fenomene, pot fi sesizate mai uşor punctele critice din cadrul dezvoltării, se evidenţiază tendinţa centrală de dezvoltare a fenomenelor, variaţiile sezoniere, ciclice şi cele întâmplătoare. Reprezentările grafice sunt utile şi pentru popularizarea faptelor şi evenimentelor deosebite din viaţa socială şi economică, ca şi pentru efectuarea diferitelor calcule.

Pentru reprezentarea grafică a seriilor dinamice se utilizează mai des în practică: historiogramele, diagramele polare şi diagramele în spirală.

5.3.2.1. Historiogramele

Historiogramele, numite şi *cronograme*, se bazează pe sistemul de coordonate rectangulare, în care pe abscisă se reprezintă timpul (caracteristica), iar pe ordonată valorile fenomenului. Sistemul de axe rectangulare oferă posibilitatea de a reprezenta termenii seriilor dinamice $(t_i; y_i)$ prin puncte în plan. În practică se utilizează numai primul cadran, deoarece termenii seriilor dinamice sunt, de regulă, mărimi pozitive.

La stabilirea scării pe cele două axe trebuie, să se respecte principiul proporţionalităţii, deoarece raportul dintre scări are o importanţă hotărâtoare asupra formei evoluţiei şi a înclinaţiei acesteia faţă de axe.

În cazul reprezentării comparabile a mai multor serii dinamice, pe acelaşi grafic, fiecare traiectorie trebuie să se deosebească clar de celelalte. În acest scop curbele (sau liniile frânte) sunt colorate sau construite diferit (linii continue, linii întrerupte, linii punctate, linii de grosimi diferite etc.).

Deşi, în aparenţă, între historiograme n-ar exista diferenţe esenţiale, la construirea lor trebuie să se ţină seama de natura seriei dinamice, respectiv, dacă reprezintă o serie dinamică de momente sau de intervale.

Deoarece seriile dinamice de momente indică nivelul atins de către fenomenul studiat în anumite momente, ale dezvoltării lui în timp, fiecare punct al liniei frânte (cronograma liniară) corespunde unei anumite valori observate la un moment dat. Punctele empirice, care se unesc pentru a obţine linia frântă empirică, sunt plasate chiar pe ordonatele ridicate de pe axa absciselor din dreptul gradaţiilor respective (fig. 5.18).

Scanned with OKEN Scanner

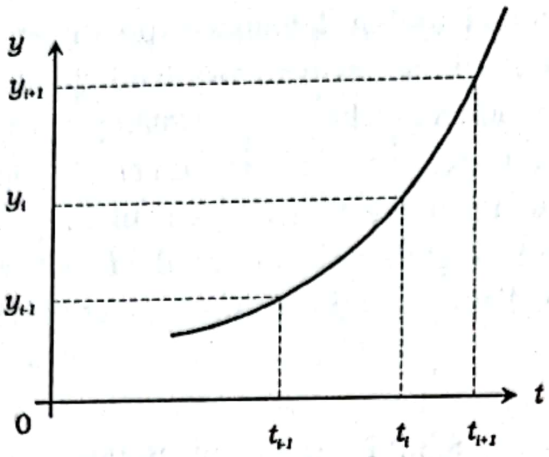


Figura 5.17. Reprezentarea grafică a seriilor dinamice de momente

Cronograma liniară a seriei din tabelul 5.23 se prezintă în figura 5.18.

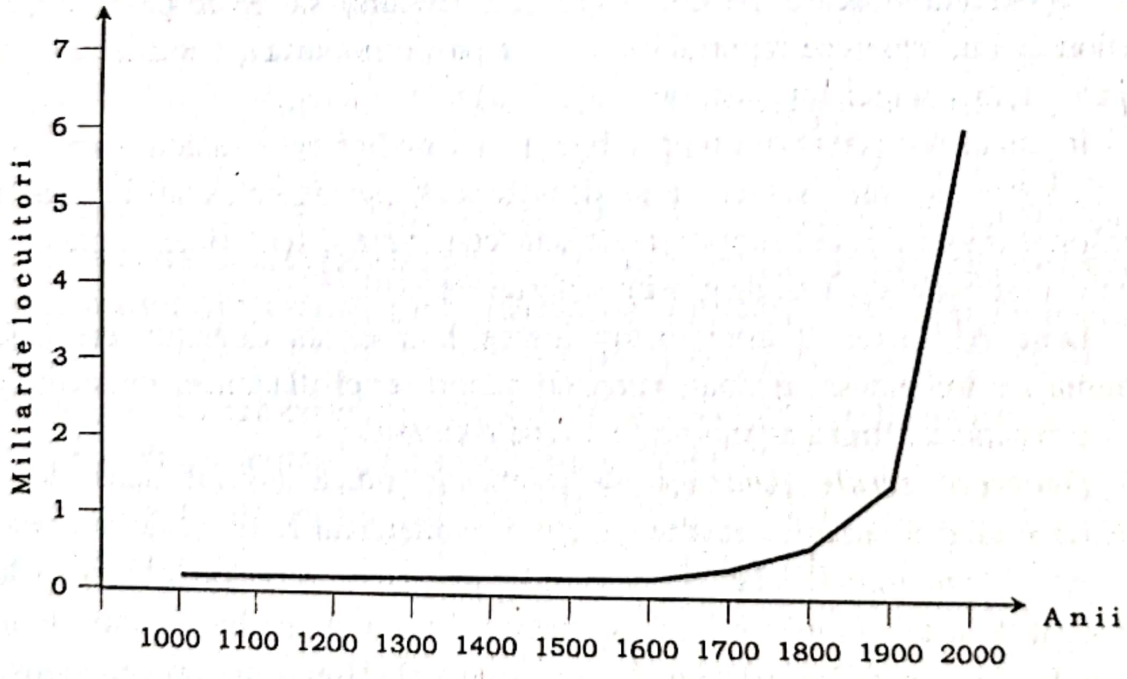


Figura 5.18. Evoluţia populaţiei globului în perioada 1000-1995 şi prognoza pentru anul 2000

Scanned with OKEN Scanner

Pe acelaşi grafic pot fi reprezentate una sau mai multe serii, ai căror termeni pot fi exprimaţi în aceeaşi unitate de măsură sau în unităţi de măsură diferite.

Dacă indicatorii tuturor seriilor, care urmează a fi reprezentate pe acelaşi grafic, sunt exprimaţi în aceeaşi unitate de măsură, se construieşte o singură scară pe axa ordonatelor. Reprezentarea pe acelaşi grafic a celor două fenomene interdependente oferă posibilitatea comparării şi analizei complexe a variaţiei lor.

În cazul în care indicatorii seriilor dinamice reprezentate sunt exprimaţi în diferite unităţi de măsură, atunci se construiesc două scări verticale (una pe partea stângă şi alta pe partea dreaptă a reţelei). Pentru a se evita în practică construcţia a două scări pe ordonată, exprimate în unităţi de măsură diferite, care complică construcţia şi analiza graficului, datele celor două serii se pot transforma în mărimi relative ale dinamicii cu acelaşi an considerat ca bază de raportare.

Dacă mărimea indicatorilor seriilor dinamice nu începe de la zero - originea axelor de coordonate - ci de la o valoare mult mai mare, se recomandă să se facă o *întrerupere a ordonatei graficului*, simbolizată printr-un zig-zag pe această axă şi să se înceapă gradarea scării deasupra întreruperii, începând cu valoarea rotundă apropiată de nivelul minim al fenomenului prezentat, dar totdeauna mai mică.

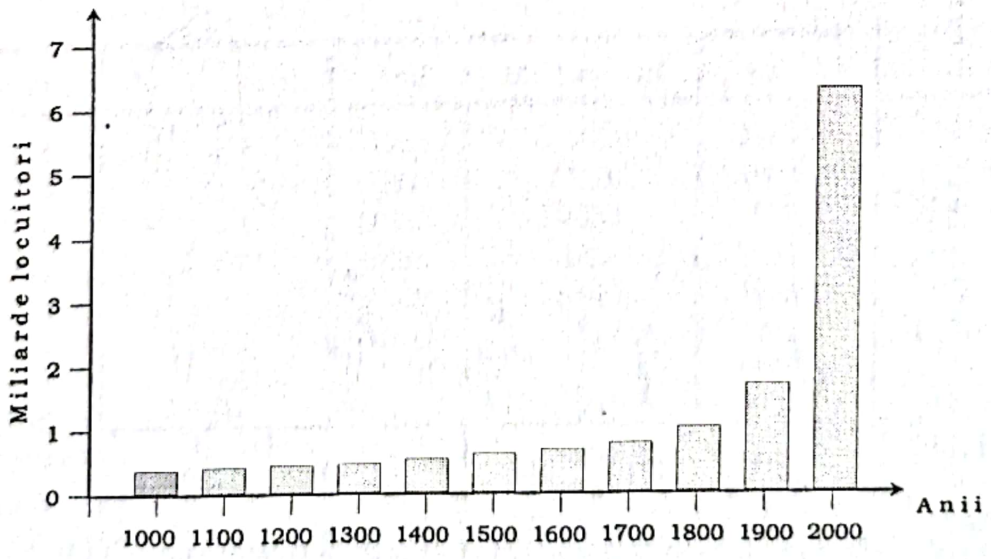


Figura 5.19. Evoluţia populaţiei lumii în perioada 1000-1995 şi prognoza pentru anul 2000

Scanned with OKEN Scanner

Pentru seriile dinamice de momente, *cronogramele* se pot construi şi *cu ajutorul coloanelor*. În acest caz din fiecare moment situat pe axa absciselor, se ridică câte o coloană, cu baza pe această axă, a cărei înălţime variază direct proporţional cu mărimea indicatorilor seriei dinamice. *Cronograma cu coloane* pune accentul pe evidenţierea mai clară a variaţiei *volumului fenomenelor*, în timp ce cronograma liniară relevă mai pregnant *tendinţa lor centrală de dezvoltare.*

Cronograma seriei dinamice de momente, din tabelul 5.23, se prezintă în figura 5.19.

Pentru a nu denatura reprezentarea proporţională a evoluţiei volumului fenomenelor în timp, este necesar ca baza coloanelor să fie de aceeaşi mărime, pentru toate momentele, iar distanţa dintre coloane să fie, în general, proporţională cu distanţa dintre momente.

Seria dinamică de momente, de mai sus, se poate ilustra grafic şi cu ajutorul unor *elemente figurative* care să reprezinte chiar populaţia. În acest caz, figurile naturale respective sunt proporţionale cu datele pe care le reprezintă (fig. 5.20).

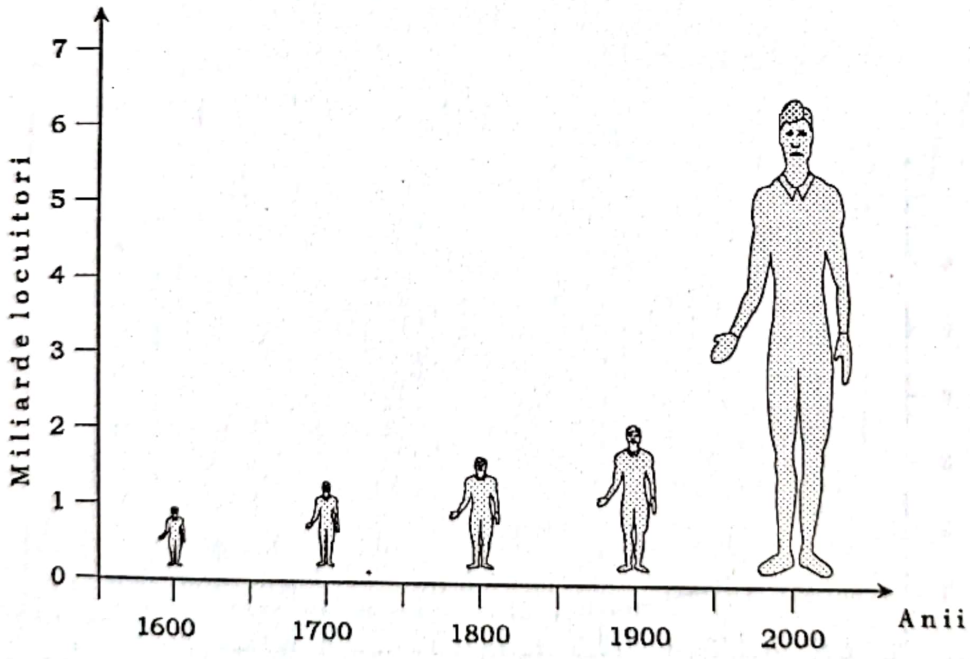


Figura 5.20. Evoluţia populaţiei lumii în perioada 1600-2000

În cazul seriilor dinamice de intervale, punctele care marchează mărimea indicatorilor corespund mijloacelor intervalelor de pe scara timpului, iar mărimea perioadei se înscrie pe această scară (fig.5.21).

Scanned with OKEN Scanner

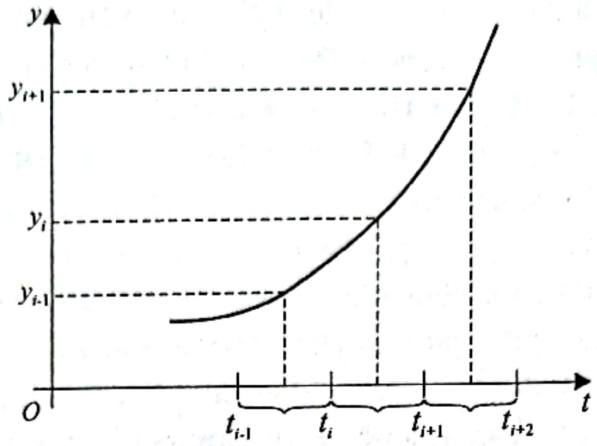


Figura 5.21. Reprezentarea grafică a seriilor de intervale

Acest procedeu se utilizează atât în cazul seriilor dinamice de intervale cu indicatorii exprimaţi în mărimi absolute, cât şi pentru cele construite din mărimi relative sau din mărimi medii.

Pentru exemplificarea construcţiei historiogramei, în cazul seriilor dinamice de intervale, considerăm seria din tabelul 5.28.

Tabelul 5.28. Producţia de oţel a *României* în anul 1938 şi în perioada 1950-1995

Anii	Mii tone oţel	Anii	Mii tone oţel	Anii	Mii tone oţel
1938	284	1965	3426	1981	13025
1950	555	1966	3670	1982	13055
1951	644	1967	4088	1983	12593
1952	698	1968	4751	1984	14437
1953	717	1969	5540	1985	13795
1954	628	1970	6517	1986	14276
1955	766	1971	6803	1987	13885
1956	779	1972	7401	1988	14314
1957	864	1973	8161	1989	14411
1958	934	1974	8848	1990	9761
1959	1420	1975	9549	1991	7130
1960	1806	1976	10733	1992	5376
1961	2126	1977	11457	1993	5446
1962	2451	1978	11779	1994	5800
1963	2704	1979	12909	1995	6557
1964	3039	1980	13175		

Sursa: *Anuarul Statistic al României 1993*, pp.506-507 şi 1996, pp. 508-509

Scanned with OKEN Scanner

Figura 5.22 arată că producţia de oţel a ţării noastre a crescut în mod continuu în cea mai mare parte a perioadei 1950 -1995 cu unele excepţii. Creşterea producţiei de oţel s-a produs mai lent la începutul perioadei şi apoi din ce în ce mai accelerat până în anul 1984, înregistrându-se o tendinţă generală ce se apropie de forma unei curbe exponenţiale, cu un început de scădere accentuată în 1990.

Din tabelul 5.28 se observă că lipsesc datele pentru perioada 1939-1949. Când seria este întreruptă pe o perioadă oarecare de timp, pentru a respecta principiul potrivit căruia unor perioade de timp egale trebuie să le corespundă pe abscisă segmente de dreaptă egale, se poate proceda astfel: anul 1938, fiind mult distanţat de perioada de după 1950, va fi separat de această perioadă fie printr-o întrerupere a graficului pe axa absciselor, fie printr-o linie verticală mai îngroşată decât liniatura reţelei. Pentru a evidenţia în grafic caracterul de continuitate al fenomenului şi pentru perioada în care seria este întreruptă, se poate uni nivelul anului 1938 cu nivelul anului 1950 printr-o linie punctată.

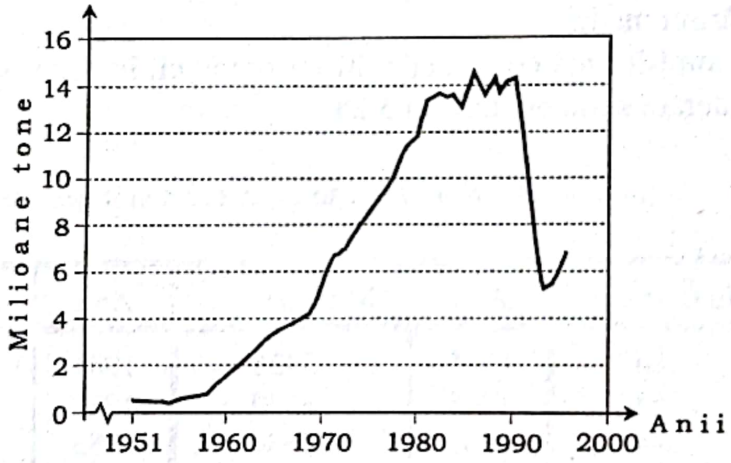


Figura 5.22. Producţia de oţel a *României* în perioada 1950-1995

Pe lângă *diagramele pe scară aritmetică*, în statistică se utilizează şi diagrame pe scară semilogaritmică şi dublulogaritmică.

Am văzut, în paragraful anterior, că diagramele cu scară uniformă sau aritmetică se caracterizează prin faptul că unor creşteri absolute egale ale unui anumit indicator le corespund segmente de dreaptă egale pe cele două axe de coordonate.

Particularitatea *diagramelor semilogaritmice* constă în faptul că pe ordonată scara este logaritmică iar pe abscisă scara timpului este aritmetică. Distanţele de pe scara ordonatelor (scara logaritmică) sunt proporţionale cu

Scanned with OKEN Scanner

logaritmii indicatorilor seriilor dinamice. Diagramele semilogaritmice, numite şi *diagrame aritlogaritmice* se utilizează în practică pentru a evidenţia proporţiile sau rapoartele procentuale şi pentru reprezentarea grafică a seriilor dinamice ai căror termeni prezintă variaţii foarte mari în timp. Diagrama semilogaritmică este utilă şi pentru compararea seriilor dinamice care se deosebesc în mare măsură între ele, din punct de vedere al nivelului de dezvoltare. Într-un astfel de grafic, variaţiile relative egale se reprezintă prin distanţe egale pe scara ordonatelor, în timp ce în diagrama cu scară aritmetică segmentele egale de pe axa ordonatelor corespund unor variaţii absolute egale ale indicatorilor.

Utilizarea scării logaritmice sau a *scării proporţiilor* se justifică în practică prin faptul că valoarea unei anumite schimbări absolute depinde, în general, de mărimea bazei în raport cu care se calculează schimbarea. De exemplu, creşterea cu 100 de unităţi faţă de o bază egală cu 100 are aproximativ aceeaşi valoare ca şi creşterea cu 1000 de unităţi la o bază egală cu 1000, deoarece în ambele cazuri creşterea relativă este egală cu 100%.

Pentru construirea scării logaritmice pe axa ordonatelor sunt necesari numai logaritmii numerelor de la 1 - 10, întrucât pentru celelalte clase de numere mantisele sunt aceleaşi, diferă numai caracteristicile. De aceea, liniatura pentru numerele 1-10 se repetă pentru numerele 10 -100,100 - 1000,1000 - 10000 ş.a.m.d.

Diagramele semilogaritmice prezintă dezavantajul că nu ne dau posibilitatea de a percepe direct şi intuitiv mărimile absolute şi diferenţele dintre ele.

5.3.2.2. Diagramele polare

Diagramele polare sau *radiale*, numite şi diagrame *în spirală*, se construiesc cu ajutorul reţelelor radiale sau curbilinii. În aceste diagrame circumferinţa cercului serveşte, de obicei, pentru notarea timpului (luni, zile, ore etc.) şi joacă rolul de abscisă în cadrul graficelor, iar ordonata se construieşte pe raza orizontală din dreapta sau pe cea verticală de sus. Mărimea razei cercului este proporţională cu nivelul mediu al indicatorilor seriei luate în studiu pe un ciclu, de regulă ciclul cu valorile cele mai mari.

Aceste diagrame se utilizează în statistică, în special, pentru reprezentarea seriilor cronologice care prezintă oscilaţii sau variaţii sezoniere ciclice, iar în cadrul unui ciclu sunt diferenţe mari de la o perioadă la alta, de exemplu: temperatura medie pe lunile unui an, volumul vânzărilor unui produs în decursul unei

Scanned with OKEN Scanner

săptămâni de zile, numărul de consumatori dintr-un restaurant pe orele unei zile etc.

Diagramele polare se pot construi astfel:

a) se ia un cerc a cărui rază este proporţională cu mărimea medie a indicatorului reprezentat;

b) cercul se împarte în atâtea părţi egale câţi indicatori ai seriei sunt cuprinşi într-un ciclu de variaţie şi prin aceste diviziuni se trasează razele cercului care împart suprafaţa lui în sectoare de cerc egale;

c) Scara graficului se construieşte pe raza orizontală din dreapta, sau pe cea verticală de sus.

În practică există două modalităţi de stabilire a originii pe scara ordonatelor (pe rază) şi anume:

- originea axei poate coincide cu centrul cercului (fig.5.23*a*);
- originea axei se consideră la intersecţia razei cu circumferinţa cercului, iar mărimea indicatorilor se măsoară fie de la circumferinţă pe prelungirea razelor în exteriorul cercului (fig. 5.23*b*), fie de la circumferinţă spre centrul cercului (fig. 5.23*c*).

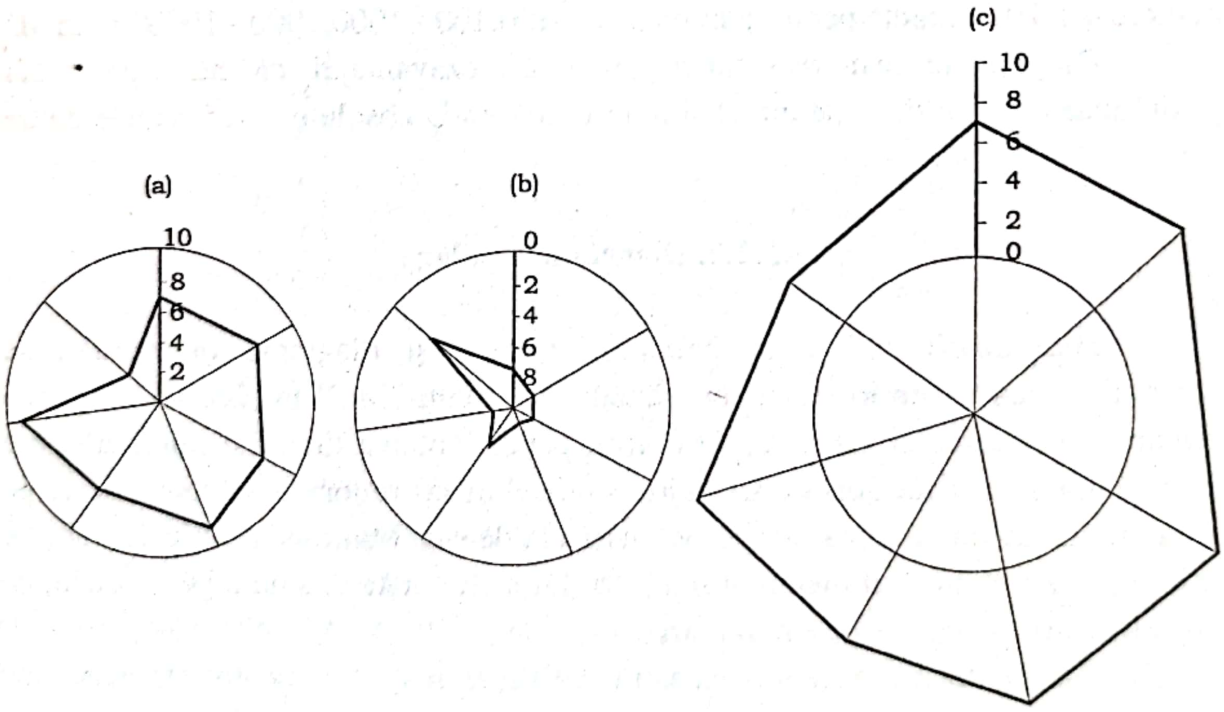


Figura 5.23. Volumul vânzărilor produsului X în decursul unei săptămâni

Scanned with OKEN Scanner

În primul caz, care este şi cel mai des utilizat în practică, dacă valoarea fenomenului de reprezentat este mai mică decât media lui, segmentul prin care se reprezintă indicatorul respectiv al seriei dinamice, măsurat de la origine spre exterior, va fi şi el mai mic decât raza, iar dacă mărimea indicatorului va fi mai mare decât raza, segmentul de dreaptă respectiv se va reprezenta şi pe prelungirea razei în exteriorul cercului.

d) După ce s-au determinat segmentele de pe raze, se unesc capetele acestora fie prin segmente de dreaptă, fie prin arce de cerc.

În cazul în care indicatorii sunt reprezentaţi prin *segmente de dreaptă*, timpul se notează în dreptul razelor, iar când indicatorii seriei dinamice sunt reprezentaţi prin *sectoare de cerc*, timpul se notează în dreptul sectoarelor de cerc.

Diagrama polară prin sectoare de cerc este foarte sugestivă, dar prezintă dezavantajul că este mai puţin exactă, deoarece suprafaţa sectoarelor de cerc creşte mai rapid decât raza acestora.

Să considerăm, pentru exemplificare, o serie dinamică care se referă la volumul vânzărilor unui produs de larg consum către populaţie, în zilele unei săptămâni (tabelul 5.29).

Tabelul 5.29. Vânzarea produsului X în zilele săptămânii [x)]

Ziua	Luni	Marţi	Miercuri	Joi	Vineri	Sâmbătă	Duminică
Mii tone	7	8	8	9	6	9	3

[x)] Date convenţionale

Seria din tabelul 5.29 se poate reprezenta grafic sub forma unei diagrame polare prin sectoare de cerc. Pentru aceasta considerăm un cerc de rază egală cu şapte unităţi, deoarece nivelul mediu al indicatorului reprezentat se apropie de această valoare. Împărţim apoi cercul în şapte sectoare egale; pe fiecare rază vom determina segmente de dreaptă a căror lungime este proporţională cu mărimea indicatorilor zilnici şi apoi se trasează prin punctele, astfel obţinute, câte un arc de cerc pentru fiecare sector (fig. 5.24).

Diagramele polare pot ilustra şi variaţia unui fenomen în decursul mai multor cicluri, când fenomenul reprezentat înregistrează o variaţie continuă şi în acelaşi sens. Se obţine astfel o diagramă polară, care ia forma unei spirale.

Scanned with OKEN Scanner

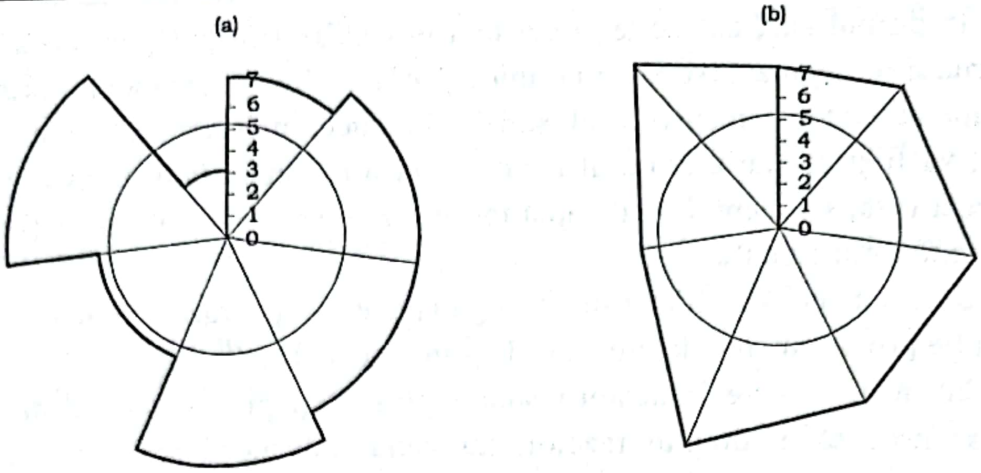


Figura 5.24. Volumul vânzărilor produsului X pe zilele unei săptămâni

Presupunem că se cunosc producţiile lunare ale unei maşini din doi ani calendaristici (tabelul 5.30).

Tabelul 5.30. Producţia produsului X

Anul \ Luna	I	F	M	A	M	I	I	A	S	O	N	D
t	9	7	13	11	12	12	8	8	15	12	12	16
$t+1$	17	13	17	14	15	13	15	15	16	24	17	20

Date convenţionale

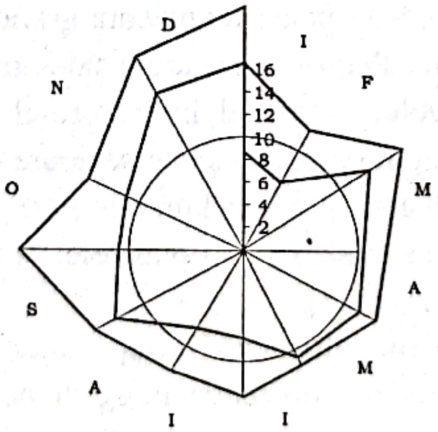


Figura 5.25. Evoluţia producţiei produsului Y în anii t şi $t+1$

Scanned with OKEN Scanner

Pentru construcţia *diagramei în spirală*, considerăm un cerc a cărui rază este proporţională cu nivelul maxim al indicatorilor din cadrul seriei dinamice. Împărţim apoi cercul în douăsprezece sectoare egale fiecare cu 30°, corespunzătoare lunilor unui an. Se alege apoi o unitate de măsură convenabilă, pentru determinarea unor segmente proporţionale cu nivelul indicatorului din luna respectivă, pe razele cercului, obţinându-se astfel mai multe puncte care, prin unire în mod consecutiv cu ajutorul unor segmente de dreaptă, formează o linie frântă, care se apropie de *forma unei spirale* (fig. 5.25).

5.4. Seriile de spaţiu

5.4.1. Noţiunea şi importanţa lor

Seriile de spaţiu sau *teritoriale* sunt acele serii statistice în care criteriul de diferenţiere a termenilor lor este teritoriul sau spaţiul (comuna, judeţul, ţara, continentul etc.) în care fenomenele au apărut, au existat sau au suferit un anumit eveniment. De exemplu: distribuţia pe judeţe a numărului şi densităţii populaţiei, a numărului salariaţilor, a principalilor indicatori ai industriei, a producţiei totale de fructe, a volumului vânzărilor de mărfuri cu amănuntul, populaţia şi densitatea ei pe continente, pe ţări, durata medie a vieţii pe ţări etc.

Înregistrarea elementelor unei colectivităţi pe diferite unităţi teritoriale se efectuează fie pentru o anumită perioadă de timp sau la un moment dat, fie în funcţie de natura fenomenelor cercetate şi acest fapt trebuie precizat în titlul tabelului care prezintă seria.

5.4.2. Reprezentarea grafică a seriilor de spaţiu

Distribuţia spaţială a fenomenelor se poate reprezenta grafic cu ajutorul *diagramelor prin benzi*, care scot în evidenţă variaţia volumului fenomenului în teritorii diferite.

Pentru a ilustra modul de construire a unui asemenea grafic, se consideră seriile de spaţiu din tabelul 5.31.

Din tabelul 5.31 rezultă că aproape două treimi din comerţul exterior al *României* în anul 1990 se efectua cu cele 10 ţări cuprinse în tabel. Ponderea cea mai importantă în comerţul exterior al ţării noastre o deţinea fosta *URSS* cu 24,2%, iar în cadrul relaţiilor comerciale exportul era mai mic decât importul cu 10,3%.

Scanned with OKEN Scanner

Dintre celelalte ţări, relaţiile comerciale cele mai dezvoltate le aveam cu Germania, cu o pondere de 11% şi volumul exportului era puţin mai mic decât volumul importului.

Structura comerţului exterior al *României* cu aceste ţări, atât la import cât şi la export, se poate reprezenta grafic cu ajutorul unei diagrame prin benzi (fig. 5.26).

Benzile au aceeaşi grosime pentru toate ţările, iar lungimea lor este proporţională cu valoarea absolută a indicatorilor reprezentaţi (valoarea exportului, respectiv valoarea importului). De fapt, în figura 5.26 sunt prezentate în paralel două diagrame, una pentru export şi cealaltă pentru import. Între cele două diagrame s-a înscris ţara, iar scara este construită în partea de sus şi cea de jos a graficului.

Tabelul 5.31. Primii zece parteneri de comerţ exterior ai *României* în anul 1990

Ţara	Total		Export		Import	
	Milioane lei	%	Milioane lei	%	Milioane lei	%
URSS	83589,9	24,2	34102,9	25,2	49487,0	23,6
Germania	38900,4	11,3	14883,3	11,0	24017,1	11,4
Arabia Saudită	17815,5	5,2	465,5	0,3	17350,0	8,3
SUA	17482,4	5,1	7875,1	5,8	9607,3	4,6
Italia	14464,3	4,2	11963,4	8,9	2500,9	1,2
Iran	14118,4	4,1	1798,9	1,3	12319,5	5,9
Polonia	10876,2	3,2	1913,3	1,4	8962,9	4,3
Egipt	9497,6	2,8	1860,9	1,4	7636,7	3,6
China	8911,4	2,6	3541,6	2,6	5369,8	2,6
Franţa	8657,5	2,5	4643,7	3,4	4013,8	1,9
Total (10 ţări)	224313,6	65,2	83048,6	61,3	150227,9	71,6
Total general	345103,2	100	135191,4	100	209911,8	100

Sursa: *Anuarul Statistic al României 1991*, CNS, Bucureşti, p.600-617

Caracterul de distribuţie teritorială a seriilor de spaţiu nu poate fi scos în evidenţă, în mod clar, numai cu ajutorul graficelor obişnuite (diagrame), ci este necesar, pentru localizarea fenomenelor cercetate, să se apeleze la un tip special de grafice, construite cu ajutorul hărţilor de contur a diviziunilor teritoriale.

Desigur, economistul nu este geograf şi reprezentările sale spaţiale nu vizează acelaşi scop; în loc de a încerca reproducerea detaliată a fenomenelor în spaţiu, economistul consideră că introduce în lucrările sale o variabilă

Scanned with OKEN Scanner

suplimentară, spaţiul. Reprezentarea grafică cu ajutorul hărţilor este tot atât de importantă pentru analiza statistico-economică şi pentru testarea ipotezelor statistice, ca şi diagramele. Scopul statisticii este şi *deglobalizarea* spaţială a colectivităţilor, relevarea diversităţii care se ignoră de către volumul acestora sau de către nivelul mediu al diferitelor caracteristici. Spaţiul pentru economist apare deci mult mai abstract decât pentru geograf.

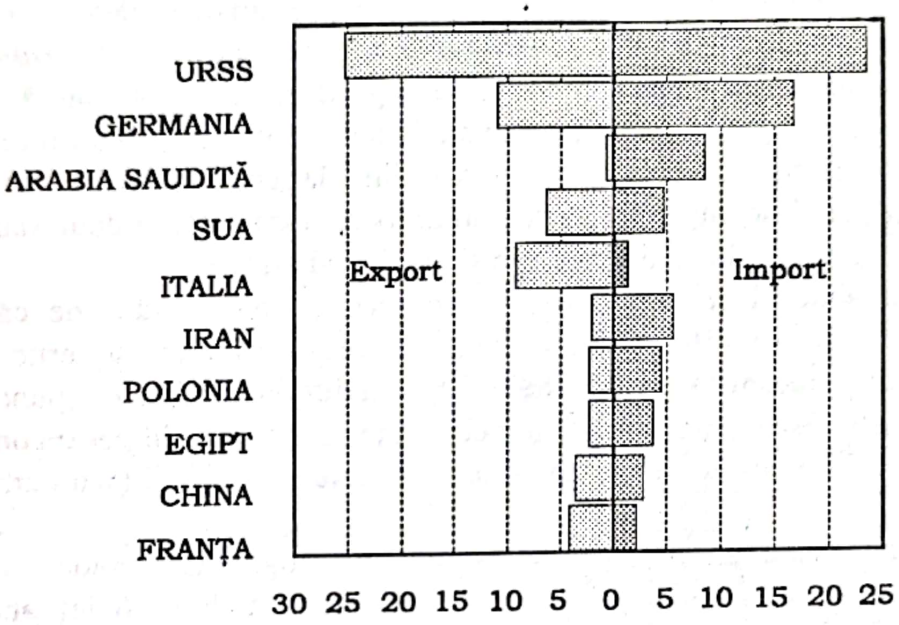


Figura 5.26. Comerţul exterior al *României* cu primele zece ţări în anul 1990

Înainte de a prezenta unele modele ale diagramelor spaţiale, se impun câteva precizări:

a) hărţile geografice sunt reprezentări prin suprafeţe: fiecare unitate geografică este reprezentată printr-o suprafaţă proporţională cu cea reală;

b) importanţa fiecărei zone geografice este foarte diferită din multe puncte de vedere;

c) subzonele unui anumit teritoriu au o individualitate proprie: caracteristicile geografice (temperatură, precipitaţii); sociodemografice (regimul de activitate, nivelul mortalităţii generale şi specifice, numărul de locuinţe etc); economice (mărimea şi forma suprafeţei, dezvoltarea industriei şi structura ei, preţurile produselor de sezon etc);

d) analiza unei cartograme trebuie să vizeze cel puţin două aspecte:

Scanned with OKEN Scanner

• *intensitatea luminoasă* trebuie să fie în concordanţă cu nivelul caracteristicii zonei luate în studiu (două zone geografice apropiate sau chiar identice se reprezintă prin aceeaşi tonalitate luminoasă, culoare sau haşurare);

• *numărul de tonalităţi* diferite care se pot utiliza într-o cartodiagramă este limitat de posibilităţile de percepţie ale omului;

Pornind de la aceste precizări de ordin general, se prezintă în continuare câteva modele de grafice ale distribuţiilor spaţiale.

Graficele specifice utilizate pentru reprezentarea distribuţiilor după o caracteristică teritorială sunt de două tipuri: *cartograme* şi *cartodiagrame*.

Cartogramele, la fel ca oricare alte grafice, trebuie să fie cât mai simple pentru a putea fi uşor de citit şi interpetat,. Uneori, din dorinţa unor reprezentări cât mai precise şi mai complete, se pierde din claritate. În astfel de cazuri este preferabil să se descompună o cartogramă, foarte încărcată, în două sau mai multe cartograme mai simple, mai expresive şi, deci, mai eficace.

Cu ajutorul *cartogramelor* se reprezintă nivelul atins de către diferite fenomene în teritorii diferite sau variaţia acestora de la o unitate teritorială la alta, cu scopul de a înlesni caracterizarea statistică a diferitelor zone din punct de vedere al caracteristicii sau a caracteristicilor cercetate. Cartogramele pot fi construite prin puncte, în culori sau în nuanţe de culori, haşurate, cu izolinii (sau curbe de nivel) etc.

Harta care stă la baza construcţiei cartogramei trebuie să aibă bine conturate unităţile teritoriale, fără să se precizeze formele de relief, apele, oraşele etc. decât în măsura în care este nevoie de localizarea fenomenelor cercetate, pentru a nu încărca în mod inutil cartograma.

Pentru reprezentarea pe hartă a variaţiei diferitelor fenomene sau a intensităţii acestora în teritorii diferite, se utilizează în practică *cartograma prin puncte*. În acest caz, indicatorii se reprezintă printr-un anumit număr de puncte, fiecare având o anumită mărime. Problema principală a construcţiei acestor cartograme constă în alegerea raţională a mărimii etalon a punctului (cerculeţului). Dacă punctele vor reprezenta valori prea mari, ele se vor uni cu cele învecinate şi vor da o nuanţă de gol teritoriului din imediata apropiere. Dimpotrivă, dacă punctele vor reprezenta valori prea mici, ele vor fi prea dese pe hartă, se vor contopi şi nu vor da o imagine clară. Cartogramele prin puncte se utilizează mai ales pentru reprezentarea seriilor formate din mărimi absolute (populaţie, producţie, vânzări etc.).

Mai răspândite în practică sunt *cartogramele cu culori diferite* şi cele cu *intensităţi diferite* ale aceleiaşi culori. Dacă se întrebuinţează culori diferite, apare dificultatea reprezentării culorilor într-o progresie optică cu scopul de a permite identificarea precisă şi logică a variaţiei şi intensităţii fenomenelor. În literatura de specialitate s-a propus următoarea scară de culori, în ordine crescătoare a

Scanned with OKEN Scanner

intensităţii: alb, galben-deschis, galben-auriu, portocaliu, roz, maro-deschis, maro-închis, verde-închis, roşu-închis şi negru. Din practică, s-a constatat că în locul utilizării mai multor culori, este preferabilă utilizarea mai multor nuanţe ale aceleiaşi culori. Intensitatea culorii este strâns legată de intensitatea fenomenului în zonele respective, astfel încât, acolo unde aceasta este mai mare, şi nuanţa culorii va fi mai pronunţată şi invers.

Cartogramele haşurate pot înlocui cartogramele în culori. Haşurarea diferitelor zone se face în baza aceloraşi principii ca şi în cazul cartogramelor în culori: cu cât intensitatea fenomenului este mai mare, cu atât haşurarea are un fond mai închis, ce se realizează fie prin desimea liniilor, fie prin grosimea lor şi invers (fig. 5.27).

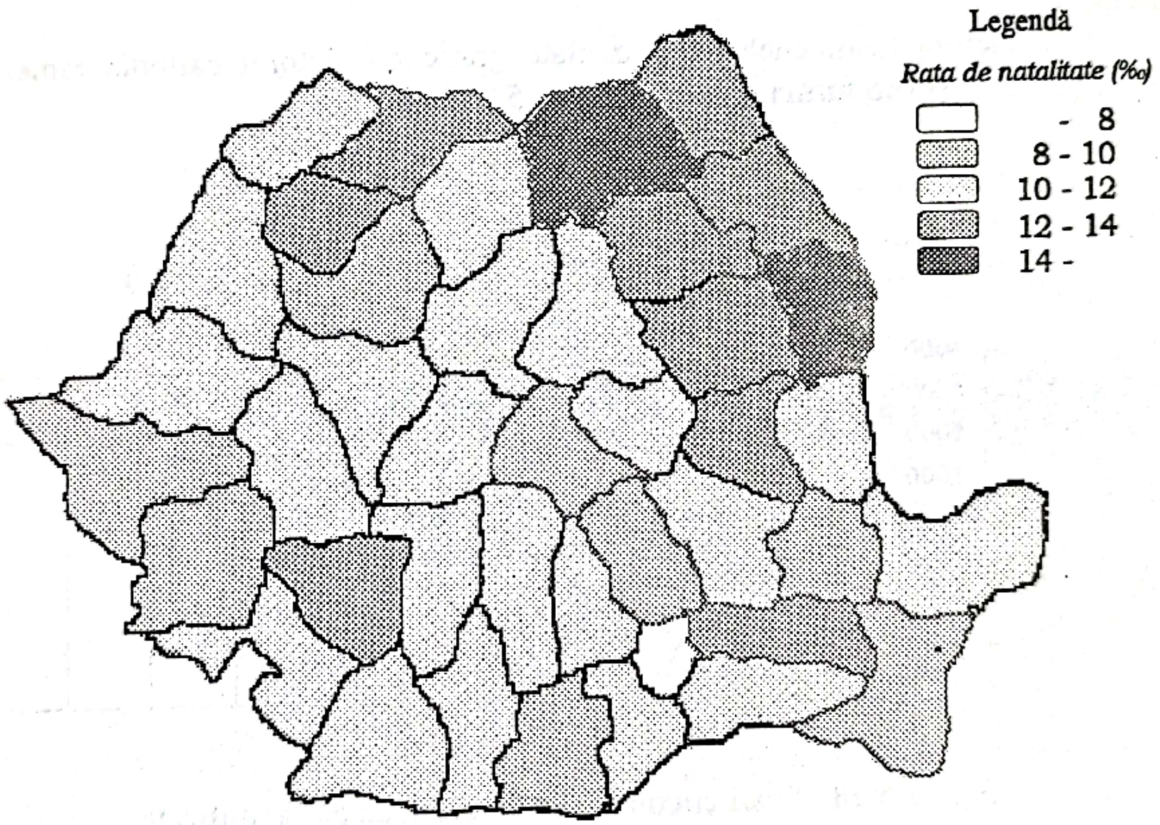


Figura 5.27. Distribuţia judeţelor *României* după rata de natalitate în anul 1994

Cartogramele cu izolinii sunt formate din mai multe curbe de nivel mai mult sau mai puţin apropiate între ele, după cum intensitatea fenomenului reprezentat este mai mare sau mai mică.

Scanned with OKEN Scanner

Cartodiagramele reprezintă o formă combinată a cartogramelor cu diagramele. Construcţia cartodiagramelor începe cu desenarea unei hărţi administrative a teritoriului luat în studiu, în interiorul căreia se înscriu diferite diagrame care reprezintă indicatorii seriei teritoriale. Diagramele cele mai utilizate în practică sunt: cercurile, pătratele, coloanele, punctele etc.

Cartodiagramele prin cercuri de structură se pot construi în două feluri. Dacă interesează numai structura, cercurile desenate pe diagramă vor fi de aceeaşi mărime. Dacă, pe lângă structură, interesează şi variaţia valorilor absolute ale indicatorilor, atunci suprafaţa cercurilor va fi proporţională cu mărimea absolută a indicatorilor. În cel de-al doilea caz este necesar ca în legenda graficului să se construiască o scară sub forma unei diagrame circulare (fig. 5.28 a şi b) pe baza unei mărimi absolute convenţionale: 1000 unităţi, 2000 unităţi, 3000 unităţi, 4000 unităţi.

Structura fenomenelor reprezentate grafic cu ajutorul cartodiagramei se poate prezenta şi sub forma unui pătrat (fig. 5.28 c).

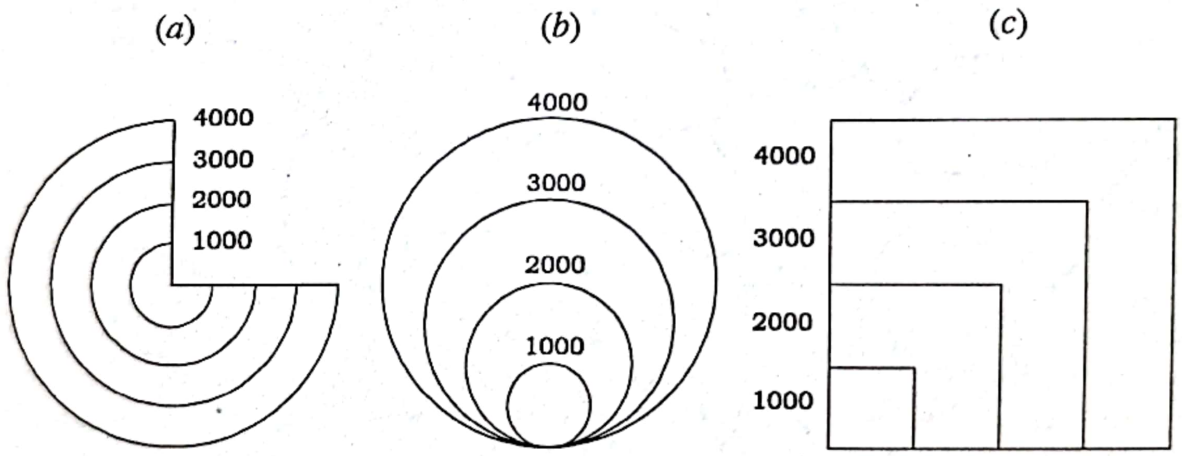


Figura 5.28. Scări circulare şi sub formă de pătrat utilizate pentru construcţia cartodiagramelor

La cartodiagramele care se construiesc cu ajutorul coloanelor, în spaţiul fiecărei zone se înscriu coloanele ce reprezintă indicatorii luaţi în calcul.

Scanned with OKEN Scanner

III. ANALIZA DISTRIBUTIILOR UNIDIMENSIONALE

Scanned with OKEN Scanner

Scanned with OKEN Scanner

6. MĂRIMILE MEDII

7. INDICATORII VARIAŢIEI ŞI CONCENTRĂRII

8. REPARTIŢIILE UNOR VARIABILE ALEATOARE

9. CARACTERISTICI ALE FORMEI DISTRIBUŢIILOR

6. MĂRIMILE MEDII

6.1. Esenţa şi rolul mărimilor medii în cunoaşterea statistică

În capitolele anterioare ale acestei lucrări am văzut cum se poate obţine o imagine generală asupra modului de distribuire a unităţilor unei colectivităţi în funcţie de o anumită caracteristică dată, pe baza seriilor statistice şi a reprezentărilor grafice. Altfel spus, am văzut cum se distribuie variantele unei variabile în cadrul unei colectivităţi pe baza informaţiilor sistematizate cu ajutorul seriilor statistice şi reprezentate vizual cu ajutorul graficelor.

Într-o cercetare concretă şi completă, această imagine generală este insuficientă. În locul analizei mulţimii variantelor caracteristicii este de preferat un indicator generalizator, care să reprezinte *nivelul general* al fenomenului studiat, *tendinţa lui centrală* de dezvoltare. Pentru a caracteriza succint o serie statistică se determină *o valoare reprezentativă*, în jurul căreia are loc o tendinţă de concentrare a nivelurilor individuale ale caracteristicii.

Fenomenele de masă din natură sau societate, supuse cercetării statistice, sunt purtătoare a mai multor caracteristici, fiecare cu un mare număr de variante, ca rezultat al îmbinării influenţelor diferiţilor factori, sub acţiunea cărora iau naştere şi se dezvoltă aceste fenomene şi care determină, totodată, diferenţierea nivelurilor caracteristicii de la o unitate statistică la alta sau de la un grup de unităţi la altul. Marea diversitate a variantelor unei caracteristici constituie o reflectare a particularităţilor individuale ale unităţilor care intră în componenţa colectivităţilor. În cadrul fenomenelor colective se regăsesc, alături de trăsăturile particulare, însuşirile generale proprii colectivităţii în ansamblul ei, ca o expresie a necesităţii interne, adică a legităţilor statistice. Variaţia unei caracteristici în cadrul unei colectivităţi este determinată de acţiunea factorilor esenţiali asociaţi cu factorii neesenţiali. *Factorii esenţiali* acţionează asupra tuturor unităţilor dintr-o colectivitate şi determină dezvoltarea lor normală sub aspectul unei anumite caracteristici sau *tendinţa normală, centrală* de dezvoltare a unei colectivităţi. Desigur, dacă aceşti factori ar acţiona asupra tuturor unităţilor colectivităţii în mod egal şi în acelaşi sens, s-ar putea ajunge la o dezvoltare uniformă, egală a tuturor unităţilor statistice sub aspectul unei caracteristici date. În realitate, asupra fenomenelor acţionează şi factori întâmplători, care determină abateri, în ambele sensuri, ale nivelurilor individuale ale unei caracteristici de la tendinţa normală de

Scanned with OKEN Scanner

dezvoltare. Deoarece influenţa lor nu poate fi eliminată, fenomenele colective prezintă o mare diversitate. În procesul complex al cunoaşterii statistice a realităţii este necesar să se separe influenţele acestor două categorii de factori. Aceasta este însă o problemă dificilă a ştiinţei în ansamblul ei care nu poate fi soluţionată separat decât de către fiecare ştiinţă particulară. Statistica, cu ajutorul metodelor sale proprii de cunoaştere a realităţii, poate efectua anumite estimaţii ale mărimii acestor influenţe, care trebuie interpretate probabilistic, în strânsă legătură cu natura fiecărei colectivităţi.

Indicatorii sintetici, generalizatori ai tendinţei centrale de dezvoltare a fenomenelor poartă denumirea de *mărimi medii.*

Esenţa mărimilor medii constă în faptul că ele caracterizează în mod sintetic, generalizat şi într-o singură expresie numerică ceea ce este esenţial, normal, tipic şi reprezentativ în cadrul unei colectivităţi statistice sub aspectul unei caracteristici date.

Media sau *valoarea tip* poate fi considerată valoarea caracteristicii corespunzătoare frecvenţei maxime sau mărimea ei situată la mijlocul şirului de valori, aranjate crescător ori descrescător, sau mărimea medie a variabilei obţinută prin calcule specifice luând în considerare toate variantele acesteia.

Mărimile medii oglindesc, deci, *influenţa factorilor esenţiali*, care au acţionat asupra unităţilor statistice dintr-o colectivitate sub aspectul unei anumite caracteristici. Influenţa factorilor care determină împrăştierea nivelurilor individuale ale unei anumite variabile, în jurul nivelului ei mediu este măsurată de către *indicatorii împrăştierii.*

Datorită trăsăturilor lor, de a reprezenta printr-o singură mărime nivelurile individuale ale caracteristicii dintr-o serie, mărimile medii îndeplinesc funcţii importante în cadrul cercetării statistice.

Cu ajutorul mediilor se pot *compara colectivităţile de fenomene* şi se pot desprinde anumite trăsături generale asupra nivelului lor de dezvoltare.

Mărimile medii au fost utilizate în practică din cele mai vechi timpuri, ele reprezentând unul din primii indicatori folosiţi de gânditorii antici pentru a obţine o valoare aproximativă, unică, din combinarea unui ansamblu de date observate. Astfel, este de presupus că astronomii şi constructorii egipteni şi cei babilonieni au utilizat mărimile medii pentru a determina acum 45 de secole poziţia soarelui şi a altor corpuri cereşti.

Într-o primă aproximare, caracterizarea distribuţiilor statistice printr-un număr unic , printr-o *valoare tip*, reduce compararea a două serii de distribuţie la a compara două valori medii. În compararea a două serii de distribuţie se întâlnesc mai multe situaţii care se pot ilustra sugestiv pe cale grafică (fig.6.1). Pentru a uşura interpretarea, vom considera cazul caracteristic al distribuţiilor unimodale, normale.

Scanned with OKEN Scanner

Două distribuţii normale de acelaşi volum, cu aceeaşi medie şi acelaşi grad de împrăştiere se prezintă în figura 6.1(a), iar în figura 6.1(b) se prezintă două distribuţii simetrice, una normală (1) şi cealaltă (2) mult aplatizată. Aceste distribuţii (fig.6.1 (b)), deşi au acelaşi nivel mediu al unei caracteristici, prezintă un grad de împrăştiere diferit: distribuţia (1) este mai omogenă decât distribuţia (2); în figura 6.1(c) se prezintă două distribuţii cu acelaşi grad de omogenitate, dar care diferă prin tendinţa lor centrală, în timp ce în figura 6.1(d) curbele, deşi au aceeaşi tendinţă centrală, diferă prin gradul şi sensul asimetriei.

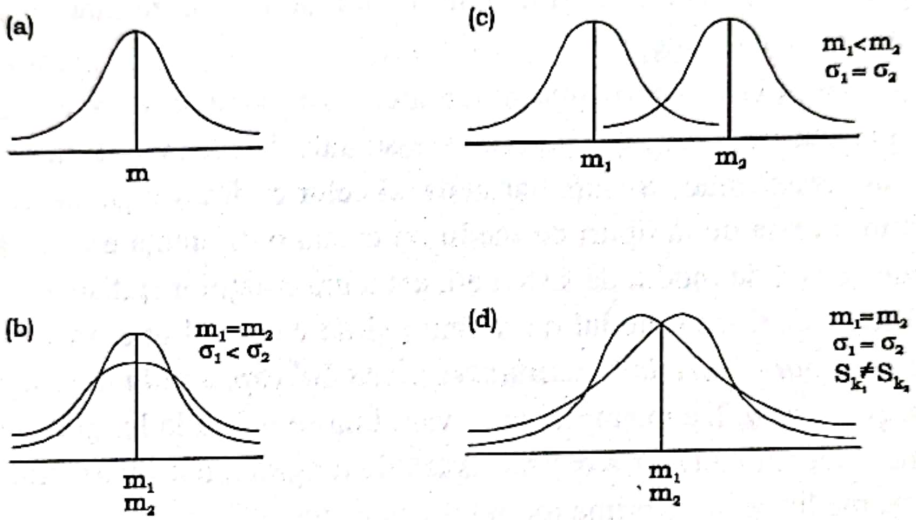


Figura 6.1. Ilustrarea geometrică a principalelor proprietăţi ale distribuţiilor

În practica statistică, mediile se utilizează pentru caracterizarea ritmului şi tendinţei de dezvoltare a fenomenelor colective, pentru studierea legăturilor dintre fenomene, pentru descoperirea legilor, legităţilor şi regularităţilor lor de dezvoltare.

Omogenitatea colectivităţilor constituie o condiţie indispensabilă pentru a *asigura mediei un conţinut real şi reprezentativ*, un *caracter ştiinţific*.

În statistica contemporană se utilizează un mare număr de mărimi medii, care diferă între ele atât prin definiţia lor, prin modul de calcul, cât şi prin conţinutul şi rolul lor specific pe care-l au în cercetarea colectivităţilor: media aritmetică, media cronologică, media pătratică, mediile mobile, media armonică, media anarmonică, media geometrică, mediile marginale, mediile condiţionate, mediile extinse, mediana, modul, mediala, medii de ordin superior etc.

Scanned with OKEN Scanner

Pentru sistematizarea studiului lor, mărimile medii se pot clasifica după criterii diferite.

a. În funcţie de *modul de determinare* a lor, mărimile medii se pot împărţi în două categorii: *indicatori propriu-zişi ai tendinţei centrale* şi *indicatori ai localizării*. Din rândul indicatorilor de localizare fac parte: mediana, modul şi mediala, iar toate celelalte sunt medii propriu-zise.

b. După importanţa lor în practica statistică, mărimile medii se pot clasifica în două categorii: *mărimi medii de bază* sau *fundamentale* şi *mărimi medii cu aplicaţii speciale*. Cele fundamentale se utilizează în mod curent în statistică şi cuprind: media aritmetică, mediana, modul şi mediala, în timp ce mediile cu aplicaţii speciale, aşa cum arată şi denumirea lor, se utilizează mai rar în practică, doar în diferite cazuri particulare.

c. O altă clasificare a mărimilor medii se poate face în funcţie de felul seriei (simplă sau cu frecvenţă) pentru care se calculează. Astfel, în cazul seriilor simple se calculează *medii simple*, iar în cazul celor cu frecvenţă, *medii ponderate*. Desigur, între aceste două tipuri de medii nu există o diferenţă esenţială, ci numai una formală, legată de modul de sistematizare a informaţiilor statistice.

Oricare ar fi însă modul de definire şi de calcul al unei mărimi medii, ea reprezintă *o valoare teoretică a caracteristicii şi se exprimă în aceleaşi unităţi de măsură ca şi aceasta*. De exemplu, dacă variabila se referă la lungimea unui obiect sau fenomen, media va fi tot o caracteristică de lungime, dacă variabila se exprimă în procente, media se va exprima tot în procente etc.

Mărimile medii, indiferent de natura lor, trebuie să satisfacă mai multe condiţii generale - cunoscute în literatura de specialitate sub denumirea de *condiţiile lui Yule*(1871 - 1951) [204, p.124-125] - după numele statisticianului englez care le-a formulat pentru prima dată în anul 1911, şi anume:

1. *Mărimile medii trebuie definite în mod precis şi obiectiv.*

2. *Orice mărime medie trebuie să ia în calcul toate nivelurile individuale ale caracteristicii* cuprinse în seria pentru care se calculează. Dacă această condiţie nu este respectată, media obţinută are un caracter aproximativ şi nu poate fi considerată ca un indicator reprezentativ.

3. *Mediile trebuie să posede o semnificaţie concretă* uşor de conceput, ele nu trebuie să aibă un caracter matematic prea abstract.

4. *Mediile trebuie să poată fi calculate cu rapiditate* şi cu uşurinţă, evitându-se pierderea de timp cu operaţii complicate de calcul. Astfel, dintre două medii care corespund în aceeaşi măsură naturii datelor iniţiale se va alege, în fiecare caz concret, aceea care se calculează mai uşor.

Scanned with OKEN Scanner

5. *Mediile trebuie să fie cât mai puţin sensibile la fluctuaţiile de selecţie*, adică pentru orice eşantion de volum n, extras dintr-o colectivitate mai numeroasă de volum N, se vor obţine în general medii diferite. Se va alege acea medie a eşantionului între care şi media colectivităţii generale există cea mai mică diferenţă. Vom vedea mai târziu sensul exact al acestei condiţii. Acum se poate totuşi sublinia faptul că gradul de sensibilitate al mediilor la fluctuaţiile de selecţie depinde de legea de distribuţie a colectivităţii totale, de felul şi volumul sondajului.

6. *Mediile trebuie să se preteze cu uşurinţă la calcule algebrice ulterioare.* Această condiţie are o importanţă mai redusă atunci când este vorba de studierea unei serii izolate, dar are un rol determinant în studiul comparativ al mai multor serii statistice.

Aşa după cum afirmă profesorul francez *H.Guitton* [88,p.128], *"mărimile medii joacă un rol foarte important în statistică şi reprezintă nu numai tendinţa centrală a unei serii, ci chiar valoarea centrală a statisticii ca ştiinţă"*. De aceea, *statistica* a fost definită de unii autori ca fiind *ştiinţa mărimilor medii* (*A.L.Bowley*[1] (1869-1957)).

6.2. Medii propriu-zise

6.2.1. Media aritmetică

Această medie este cel mai des utilizată în practica statistică deoarece este uşor de calculat şi de interpretat. Ori de câte ori se vorbeşte de medie, fără a se face vreo altă precizare cu privire la felul ei, se înţelege că este vorba de media aritmetică.

Media aritmetică a unui şir de valori individuale ale unei caracteristici este egală cu raportul dintre suma şi numărul lor.

În practică, în funcţie de natura seriei pentru care se calculează, se disting două forme diferite ale mediei aritmetice:

1) în cazul unei serii simple - în care fiecare nivel individual al unei caracteristici este purtat de o singură unitate statistică sau de un număr constant de unităţi - se calculează *media aritmetică simplă* (neponderată);

2) în cazul unei serii cu frecvenţă se calculează *media aritmetică ponderată*.

[1] A. Bowley - *Elements of Statistics*, Scribner, London, 1901.

Scanned with OKEN Scanner

Media aritmetică ponderată a fost introdusă în ştiinţă de către *Roger Cotes*[2] în anul 1712.

Dacă considerăm o variabilă X cu variantele ei $x_1, x_2, \ldots, x_i, \ldots, x_n$ atunci media aritmetică a acestei variabile, notată cu $\bar{x}$, este egală cu:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_i + \cdots + x_n}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad sau \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad (6.1)$$

Datorită faptului că fiecare nivel individual x_i al caracteristicii se întâlneşte o singură dată, în cadrul seriei luate în consideraţie, formula (6.1) poartă denumirea de *medie aritmetică simplă*.

Se vede clar că, pentru determinarea mediei aritmetice, nu este necesar, în mod obligatoriu, să cunoaştem toate nivelurile individuale ale unei caracteristici, ci doar suma nivelurilor individuale şi numărul acestor niveluri. De exemplu, pentru determinarea salariului mediu al personalului dintr-o întreprindere există două posibilităţi: fie să însumăm salariile lunare efective ale întregului personal şi apoi să raportăm rezultatul obţinut la numărul personalului, fie să raportăm direct mărimea fondului lunar de salarizare la numărul personalului. Un alt exemplu, pentru a se obţine numărul mediu de membri dintr-o familie, într-o ţară, se împarte numărul total al populaţiei ţării, rezultat în urma unui recensământ, la numărul de familii. Să considerăm acum volumul total al vânzărilor de mărfuri cu amănuntul din *România* şi numărul de locuitori în diferiţi ani (tabelul 6.1) şi să determinăm vânzările medii pe un locuitor.

Volumul total al vânzărilor de mărfuri cu amănuntul reprezintă deci numărătorul formulei mediei aritmetice $\sum_{i=1}^{n} x_i$, iar numărul populaţiei numitorul ei n. Prin efectuarea raportului se determină volumul mediu al vânzărilor de mărfuri cu amănuntul pe un locuitor (col.4 din tabelul 6.1).

---

[2] R. Cotes - *Aestimatio Errorum in Mixta Mathesi per variationes partium Trianguli plani et sphaerici*, Opera Miscellania, ed. R. Smith, Cambridge, 1722

Scanned with OKEN Scanner

Tabelul 6.1. Vânzările de mărfuri cu amănuntul în *România*

Anii	Milioane lei[x)]	Milioane locuitori la 1.VII	Vânzări medii pe un locuitor (lei)
1	2	3	4
1950	12 050	16,3	739
1960	41 507	18,4	2 256
1970	96 708	20,3	4 764
1980	213 085	22,2	9 598
1990	360 354	23,2	15 532

x) În preţurile fiecărui an

Sursa: Prelucrat după *Anuarul statistic* al *României*, 1991, CNS, Bucureşti, p.564

Dacă numărul variantelor unei variabile este foarte mare, calculul mediei aritmetice se poate simplifica formând, în prealabil, seria cu frecvenţă.

În cazul grupării nivelurilor individuale ale caracteristicii pe variantele ei, media aritmetică se va determina pe baza definiţiei formulate anterior, folosind informaţiile sistematizate cu ajutorul tabelului 6.2.

Tabelul 6.2. Distribuţia unei caracteristici pe variante şi calculul valorii ei totale

Caracteristica x_i	Frecvenţa n_i	$x_i n_i$
x_1	n_1	$x_1 n_1$
x_2	n_2	$x_2 n_2$
x_3	n_3	$x_3 n_3$
⋮	⋮	⋮
x_i	n_i	$x_i n_i$
⋮	⋮	⋮
x_k	n_k	$x_k n_k$
Total	$\sum_{i=1}^{k} n_i$	$\sum_{i=1}^{k} x_i n_i$

Scanned with OKEN Scanner

$$\bar{x} = \frac{x_1 + x_1 + \cdots + x_2 + x_2 + \cdots + x_i + x_i + \cdots + x_k}{n_1 + n_2 + \cdots + n_i + \cdots + n_k} =$$

$$= \frac{x_1 n_1 + x_2 n_2 + \cdots + x_i n_i + \cdots + x_k n_k}{n_1 + n_2 + \cdots + n_i + \cdots + n_k} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}$$

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \qquad (6.2)$$

Dacă media aritmetică se calculează pentru o serie cu frecvenţă şi cu caracteristica distribuită pe intervale (tabelul 6.3), atunci este necesar să se determine mai întâi câte o *valoare reprezentativă a fiecărui interval*. Admiţându-se convenţional că, în cadrul fiecărui interval, caracteristica se distribuie în mod liniar uniform, *mijlocul intervalului* respectiv x_i' este valoarea reprezentativă a fiecărui interval.

Tabelul 6.3. Distribuţia unei caracteristici pe intervale de variaţie

Caracteristica (x_i)	Frecvenţa absolută (n_i)	Mijlocul intervalului (x_i')	$x_i' n_i$
$x_0 - x_1$	n_1	x_1'	$x_1' n_1$
$x_1 - x_2$	n_2	x_2'	$x_2' n_2$
$x_2 - x_3$	n_3	x_3'	$x_3' n_3$
⋮	⋮	⋮	⋮
$x_{i-1} - x_i$	n_i	x_i'	$x_i' n_i$
⋮	⋮	⋮	⋮
$x_{k-1} - x_k$	n_k	x_k'	$x_k' n_k$
Total	$\sum_{i=1}^{k} n_i$	-	$\sum_{i=1}^{k} x_i' n_i$

Scanned with OKEN Scanner

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i' n_i}{\sum_{i=1}^{k} n_i} \qquad (6.3)$$

unde $$x_i' = \frac{x_{i-1} + x_i}{2} \qquad (6.4)$$

Desigur, prin admiterea acestui convenţionalism se comite o anumită eroare sistematică deoarece, pe de o parte, valorile caracteristicii dintr-un interval $x_{i-1} - x_i$ sunt înlocuite cu o singură valoare x_i', iar pe de altă parte, ipoteza distribuţiei uniforme a caracteristicii în fiecare interval este îndeplinită în realitate numai în mod excepţional. Aceste erori variază, de asemeni, proporţional cu mărimea intervalelor de grupare. Considerând în locul fiecărui interval mijlocul acestuia, distribuţia pe intervale se transformă într-o distribuţie obişnuită cu frecvenţă pe variante ale caracteristicii, pentru care se aplică metode specifice de calcul şi analiză statistică.

În cazul unor serii cu intervale, în care unul sau ambele intervale marginale sunt deschise, media aritmetică are un caracter aproximativ.

6.2.1.1. Proprietăţile mediei aritmetice

Media aritmetică are o serie de proprietăţi a căror cunoaştere este utilă pentru mai buna înţelegere a conţinutului acestui indicator, ca şi pentru efectuarea unor calcule simplificate ale ei.

1. *Media aritmetică este o valoare internă a unei serii.*

Această proprietate exprimă faptul că media aritmetică a unei variabile este cuprinsă între cea mai mică valoare a caracteristicii x_{min} şi cea mai mare x_{max}.

Să considerăm, pentru demonstrare, cazul mai general al unei serii cu frecvenţă:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_s \end{pmatrix}$$

în care $$x_1 < x_2 < x_3 < \cdots < x_i < \cdots < x_s$$

Scanned with OKEN Scanner

deci $x_1 = x_{min}$ *si* $x_s = x_{max}$

Din relaţiile de mai sus rezultă:

$$x_1 < \bar{x} < x_s$$

2. *Media aritmetică a unui şir de valori constante este egală cu valoarea lor comună.*

Dacă în seria cu frecvenţă, de mai sus, considerăm:

$$x_1 = x_2 = x_3 = \cdots = x_i = \cdots = x_s = k$$

rezultă

$$\bar{x} = \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} = \frac{\sum_{i=1}^{s} k n_i}{\sum_{i=1}^{s} n_i} = k \qquad (6.6)$$

3. *Media aritmetică este asociativă.*

Aceasta înseamnă că dacă se înlocuiesc, în cadrul unei serii statistice, mai multe niveluri individuale ale caracteristicii cu media lor, atunci media seriei modificate nu se schimbă.

Într-adevăr din seria:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_{p-1}, x_p, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_{p-1}, n_p, \ldots, n_s \end{pmatrix}$$

înlocuind primele p valori ale caracteristicii cu media lor $\bar{x}_p$, în formula mediei aritmetice ponderate, se obţine:

$$\bar{x}_p = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \cdots + x_p n_p}{n_1 + n_2 + n_3 + \cdots + n_p} = \frac{\sum_{i=1}^{p} x_i n_i}{\sum_{i=1}^{p} n_i} =$$

$$= \frac{\bar{x}_p n_1 + \bar{x}_p n_2 + \cdots + \bar{x}_p n_p + x_{p+1} n_{p+1} + \cdots + x_s n_s}{n_1 + n_2 + \cdots + n_i + \cdots + n_s} =$$

$$= \frac{\frac{\sum_{i=1}^{p} x_i n_i}{\sum_{i=1}^{p} n_i} \sum_{i=1}^{p} n_i + x_{p+1} n_{p+1} + \cdots + x_s n_s}{\sum_{i=1}^{s} n_i} = \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} = \bar{x} \qquad (6.7)$$

În particular, dacă într-o serie statistică se înlocuiesc toate nivelurile individuale ale caracteristicii cu media lor, atunci volumul total al caracteristicii rămâne neschimbat.

Astfel $$\sum_{i=1}^{s} \bar{x} n_i = \bar{x} \sum_{i=1}^{s} n_i$$

şi $$\bar{x} \sum_{i=1}^{s} n_i = \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} \sum_{i=1}^{s} n_i = \sum_{i=1}^{s} x_i n_i$$

deci $$\sum_{i=1}^{s} \bar{x} n_i = \sum_{i=1}^{s} x_i n_i \qquad (6.8)$$

4. *Media aritmetică creşte sau descreşte atunci când una sau mai multe variante ale variabilei cresc sau descresc.*

Să presupunem, pentru demonstrare, că numai valoarea x_p creşte sau scade cu o cantitate constantă q. Seria modificată va fi:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, (x_p \pm q), x_{p+1}, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_p, n_{p+1}, \ldots, n_s \end{pmatrix} \qquad q > 0$$

Media aritmetică 2 a caracteristicii seriei modificate, va fi:

Scanned with OKEN Scanner

$$\bar{x}' = \frac{x_1 n_1 + x_2 n_2 + x_3 n_3 + \cdots + (x_p \pm q) n_p + x_{p+1} n_{p+1} + \cdots + x_s n_s}{n_1 + n_2 + n_3 + \cdots + n_p + n_{p+1} + \cdots + n_s} =$$

$$= \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} \pm \frac{q n_p}{\sum_{i=1}^{s} n_i} = \bar{x} \pm \frac{q n_p}{\sum_{i=1}^{s} n_i}$$

deci
$$\bar{x}' = \bar{x} \pm \frac{q n_p}{\sum_{i=1}^{s} n_i} \qquad (6.9)$$

În cazul general considerăm că $r - p$ valori ale caracteristicii se modifică cu o constantă q, iar media seriei modificate va fi:

$$\bar{x}' = \bar{x} \pm \frac{\sum_{p}^{r} q n_i}{\sum_{i=1}^{s} n_i} = \bar{x} + q \quad \text{unde} \quad p < r < n \qquad (6.10)$$

5. *Media aritmetică este translativă.*

Aceasta înseamnă că dacă se măreşte sau se micşorează fiecare nivel individual al caracteristicii dintr-o serie, cu o cantitate constantă a, media se micşorează sau se măreşte cu acea constantă $(a \neq \bar{x})$.

În cazul unei serii simple, media seriei modificate devine:

$$\bar{x}' = \frac{\sum_{i=1}^{n} (x_i \pm a)}{n} = \frac{\sum_{i=1}^{n} x_i \pm na}{n} = \bar{x} \pm a$$

deci
$$\bar{x}' = \bar{x} \pm a \qquad (6.11)$$

Pentru o serie cu frecvenţă media seriei modificate va fi:

Scanned with OKEN Scanner

$$\bar{x}' = \frac{\sum_{i=1}^{s}(x_i \pm a)n_i}{\sum_{i=1}^{s} n_i} = \frac{\sum_{i=1}^{s} x_i n_i \pm a\sum_{i=1}^{s} n_i}{\sum_{i=1}^{s} n_i} = \bar{x} \pm a$$

Pe baza proprietăţii (5), rezultă formulele de calcul simplificat al mediei aritmetice:

- pentru o serie simplă: $\bar{x} = \frac{\sum_{i=1}^{n}(x_i \pm a)}{n} \mp a$

- pentru o serie cu frecvenţă: $\bar{x} = \frac{\sum_{i=1}^{s}(x_i \pm a)n_i}{\sum_{i=1}^{s} n_i} \mp a$

Pentru a obţine o simplificare maximă a calculelor, constanta a se alege egală cu un nivel al caracteristicii din centrul seriei ordonate, apropiat, deci, de valoarea medie $\bar{x}$.

6. *Dacă se măreşte sau se micşorează fiecare variantă a caracteristicii de un anumit număr de ori k, atunci şi media aritmetică a seriei modificate se măreşte sau se micşorează de acelaşi număr de ori k, $k \neq 0$.*

În cazul unei serii simple, media seriei transformate va fi:

$$\bar{x}' = \frac{\sum_{i=1}^{n} \frac{x_i}{k}}{n} = \frac{\bar{x}}{k}$$

sau $$\bar{x} = k\bar{x}'$$

şi $$\bar{x}'' = \frac{\sum_{i=1}^{n} x_i k}{n} = k\bar{x}$$

sau $$\bar{x} = \frac{\bar{x}''}{k}$$

În cazul unei serii cu frecvenţă, media seriei modificate devine:

$$\bar{x}' = \frac{\sum_{i=1}^{s} \frac{x_i}{k} n_i}{\sum_{i=1}^{s} n_i} = \frac{1}{k}\bar{x} \qquad \text{sau} \qquad \bar{x} = k\bar{x}'$$

şi

$$\bar{x}'' = \frac{\sum_{i=1}^{s} k x_i n_i}{\sum_{i=1}^{s} n_i} = k\bar{x} \qquad \text{sau} \qquad \bar{x} = \frac{\bar{x}''}{k}$$

În baza proprietăţii (6) se obţin următoarele formule de calcul simplificat:

- pentru o serie simplă:

$$\bar{x} = \frac{\sum_{i=1}^{n} \frac{x_i}{k}}{n} \cdot k \qquad (6.14)$$

- pentru o serie cu frecvenţă:

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i}{k} n_i}{\sum_{i=1}^{s} n_i} \cdot k \qquad (6.15)$$

În formulele (6.14) şi (6.15), pentru a se obţine simplificarea maximă a calculelor, este necesar ca valoarea constantei k să fie egală cu cel mai mare divizor comun al variantelor x_i.

7. *Dacă frecvenţele fiecărei variante a caracteristicii se măresc sau se micşorează de un anumit număr de ori m, constant, atunci media aritmetică rămâne neschimbată.*

Desigur, această proprietate se aplică numai mediilor calculate pentru seriile cu frecvenţă.

Scanned with OKEN Scanner

$$\bar{x}' = \frac{\sum_{i=1}^{s} x_i \frac{n_i}{m}}{\sum_{i=1}^{s} \frac{n_i}{m}} = \frac{\frac{1}{m}\sum_{i=1}^{s} x_i n_i}{\frac{1}{m}\sum_{i=1}^{s} n_i} = \bar{x} \qquad m \neq 0$$

Valoarea constantei m se consideră egală cu cel mai mare divizor comun al frecvenţelor pentru a se obţine, prin aplicarea acestei proprietăţi, cea mai mare simplificare a calculelor.

În particular, dacă constanta m este egală cu totalul frecvenţelor prin raportarea lui n_i la m, se face trecerea de la o serie cu frecvenţe absolute la seria corespunzătoare cu frecvenţe relative. Se ştie, de asemenea, că:

$$f_i = \frac{n_i}{n} \qquad \text{sau} \qquad f_i\% = \frac{n_i}{n} \cdot 100$$

de unde rezultă $\sum_{i=1}^{s} f_i = 1$ şi $\sum_{i=1}^{s} f_i\% = 100$

Media aritmetică a unei serii cu frecvenţe relative devine:

$$\bar{x} = \frac{\sum_{i=1}^{s} x_i \frac{n_i}{n}}{\sum_{i=1}^{s} \frac{n_i}{n}} = \frac{\sum_{i=1}^{s} x_i f_i}{\sum_{i=1}^{s} f_i} = \bar{x}$$

$$\bar{x} = \sum_{i=1}^{s} x_i f_i \qquad \text{sau} \qquad \bar{x} = \frac{\sum_{i=1}^{s} x_i f_i}{100}$$

Rezultă că media aritmetică a unei serii nu se schimbă, dacă se calculează pe baza frecvenţelor absolute sau pe baza celor relative.

Evident, dacă se utilizează, în calculul mediei aritmetice, concomitent mai multe proprietăţi ale ei, se obţine o simplificare mai mare a calculelor.

Astfel, utilizând proprietăţile 5 şi 6 se obţin următoarele *formule de calcul simplificat*:

Scanned with OKEN Scanner

- pentru o serie simplă: $$\bar{x} = \frac{\sum_{i=1}^{n} \frac{x_i - a}{k}}{n} k + a$$

- pentru o serie cu frecvenţă absolută: $$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i - a}{k} n_i}{\sum_{i=1}^{s} n_i} k + a$$

- pentru o serie cu frecvenţă relativă:

$$\bar{x} = a + k \sum_{i=1}^{s} \frac{x_i - a}{k} f_i \qquad \text{şi} \qquad \bar{x} = a + k \cdot \frac{\sum_{i=1}^{s} \frac{x_i - a}{k} f_i \%}{100}$$

Utilizând proprietăţile 5, 6 şi 7 se obţine cea mai mare simplificare a calculelor:

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i - a}{k} \cdot \frac{n_i}{m}}{\sum_{i=1}^{s} \frac{n_i}{m}} \cdot k + a \tag{6.16}$$

8. *Suma algebrică a abaterilor nivelurilor individuale ale caracteristicii de la nivelul ei mediu este egală cu zero.*

Sau, altfel spus, abaterile în plus de la medie se compensează cu abaterile în minus. Această proprietate arată că media caracteristicii este o valoare reprezentativă a seriei.

Pentru o serie simplă: $$\sum_{i=1}^{n} (x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - n \frac{\sum_{i=1}^{n} x_i}{n} = 0$$

deci $$\sum_{i=1}^{n} (x_i - \bar{x}) = 0 \tag{6.17}$$

Pentru o serie cu frecvenţă:

$$\sum_{i=1}^{s}(x_i-\bar{x})n_i=\sum_{i=1}^{s}x_in_i-\bar{x}\sum_{i=1}^{s}n_i=\sum_{i=1}^{s}x_in_i-\frac{\sum_{i=1}^{s}x_in_i}{\sum_{i=1}^{s}n_i}\sum_{i=1}^{s}n_i=0$$

deci
$$\sum_{i=1}^{s}(x_i-\bar{x})n_i=0 \qquad (6.18)$$

şi
$$\sum_{i=1}^{s}(x_i-\bar{x})f_i=\sum_{i=1}^{s}x_if_i-\bar{x}\sum_{i=1}^{s}f_i=\sum_{i=1}^{s}x_if_i-\frac{\sum_{i=1}^{s}x_if_i}{\sum_{i=1}^{s}f_i}\sum_{i=1}^{s}f_i=0$$

deci
$$\sum_{i=1}^{s}(x_i-\bar{x})f_i=0$$

9. *Suma pătratelor abaterilor nivelurilor individuale ale caracteristicii de la nivelul ei mediu este mai mică decât suma pătratelor abaterilor de la oricare alt număr* a, $a\neq\bar{x}$.

În cazul unei serii simple se verifică relaţia:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2<\sum_{i=1}^{n}(x_i-a)^2$$

Pentru demonstrarea relaţiei (6.19), considerăm al doilea membru al acestei inegalităţi, în care adunăm şi scădem constanta $\bar{x}$:

$$\sum_{i=1}^{n}(x_i-a)^2=\sum_{i=1}^{n}(x_i-\bar{x}+\bar{x}-a)^2=\sum_{i=1}^{n}\left[(x_i-\bar{x})+(\bar{x}-a)\right]^2=$$

$$=\sum_{i=1}^{n}\left[(x_i-\bar{x})^2+2(x_i-\bar{x})(\bar{x}-a)+(\bar{x}-a)^2\right]=$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + 2(\bar{x} - a) \sum_{i=1}^{n} (x_i - \bar{x}) + n(\bar{x} - a)^2 =$$

$$= \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - a)^2$$

de unde $$\sum_{i=1}^{n} (x_i - a)^2 - \sum_{i=1}^{n} (x_i - \bar{x})^2 = n(\bar{x} - a)^2$$

Rezultă că, diferenţa dintre suma pătratelor abaterilor nivelurilor individuale ale caracteristicii de la o constantă a şi suma pătratelor abaterilor nivelurilor individuale ale acesteia de la nivelul ei mediu este egală cu o anumită cantitate pozitivă şi constantă.

În cazul unei serii cu frecvenţă are loc relaţia:

$$\sum_{i=1}^{s} (x_i - \bar{x})^2 n_i < \sum_{i=1}^{s} (x_i - a)^2 n_i \qquad 0 < a \neq x \tag{6.20}$$

Demonstraţia acestei relaţii se face în mod analog cu cea din cazul seriilor simple.

10. *Media aritmetică* $\bar{x}$ *a caracteristicii unei colectivităţi împărţită în* k *grupe este egală cu media aritmetică a mediilor de grupă* $\bar{x}_j$ *ponderate cu volumul grupelor* n_j.

$$\frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \bar{x}_3 n_3 + \cdots + \bar{x}_j n_j + \cdots + \bar{x}_k n_k}{n_1 + n_2 + n_3 + \cdots + n_j + \cdots + n_k} = \frac{\sum_{j=1}^{k} \bar{x}_j n_j}{\sum_{i=1}^{k} n_j} =$$

Scanned with OKEN Scanner

Această proprietate are numeroase aplicaţii în practică. De exemplu, dacă într-o întreprindere se cunoaşte productivitatea medie a muncii pe secţii, atunci media productivităţii muncii pe întreprindere se poate determina ca o medie a productivităţilor pe secţii ponderate cu numărul personalului din fiecare secţie. De asemenea, salariul mediu lunar pe o ramură a economiei naţionale se poate determina ca o medie aritmetică ponderată a salariilor medii din întreprinderile componente ponderate cu numărul personalului din fiecare întreprindere. Media generală a studenţilor unui an de studii, la sfârşitul anului universitar, se poate calcula ca o medie a mediilor de grupă ponderate cu numărul studenţilor din fiecare grupă.

11. *Media aritmetică a sumei pentru k serii este egală cu suma mediilor tuturor seriilor.*

Să considerăm, pentru demonstrarea acestei proprietăţi, variabilele:

$$\begin{aligned}
X_1&: \ (x_{11}, x_{12}, x_{13}, \ldots, x_{1j}, \ldots, x_{1n}) \\
X_2&: \ (x_{21}, x_{22}, x_{23}, \ldots, x_{2j}, \ldots, x_{2n}) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
X_i&: \ (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ij}, \ldots, x_{in}) \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
X_k&: \ (x_{k1}, x_{k2}, x_{k3}, \ldots, x_{kj}, \ldots, x_{kn})
\end{aligned}$$

Media sumei a k variabile este egală cu:

$$\overline{X_1 + X_2 + \cdots + X_i + \cdots + X_k} = \frac{\sum_j^k x_{1j}}{n} + \frac{\sum_j^k x_{2j}}{n} + \cdots + \frac{\sum_j^k x_{ij}}{n} + \cdots + \frac{\sum_j^k x_{kj}}{n} =$$

$$\frac{n\overline{X}_1}{n} + \frac{n\overline{X}_2}{n} + \cdots + \frac{n\overline{X}_i}{n} + \cdots + \frac{n\overline{X}_k}{n} = \overline{X}_1 + \overline{X}_2 + \cdots + \overline{X}_i + \cdots + \overline{X}_k = \sum_{i=1}^{k} \overline{X}_i$$

deci $$\overline{X_1 + X_2 + \cdots + X_i + \cdots + X_k} = \sum_i \overline{X}_i \qquad (6.21)$$

În particular, pentru două variabile X şi Y, proprietatea 11 se verifică astfel:

- în cazul seriilor simple:

$$X: \quad (x_1 \quad x_2 \quad \dots \quad x_i \quad \dots \quad x_n)$$

$$Y: \quad (y_1 \quad y_2 \quad \dots \quad y_i \quad \dots \quad y_n)$$

$$Z = X + Y$$

$$Z: \quad ((x_1 + y_1); (x_2 + y_2); \quad \dots \quad ; (x_i + y_i); \quad \dots \quad ; (x_n + y_n))$$

$$\bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$$

$$\bar{z} = \frac{\sum_{i=1}^{n} (x_i + y_i)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} y_i}{n} = \bar{x} + \bar{y}$$

$$\bar{z} = \bar{x} + \bar{y}$$

- în cazul seriilor ponderate:

$$X: \quad \begin{pmatrix} x_1, & x_2, & \dots, & x_i, & \dots, & x_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

$$Y: \quad \begin{pmatrix} y_1, & y_2, & \dots, & y_i, & \dots, & y_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

$$Z = X + Y$$

$$Z: \quad \begin{pmatrix} x_1 + y_1, & x_2 + y_2, & \dots, & x_i + y_i, & \dots, & x_k + y_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

Scanned with OKEN Scanner

În particular, pentru două variabile X şi Y, proprietatea 11 se verifică astfel:

- în cazul seriilor simple:

$$X: \quad (x_1 \quad x_2 \quad \dots \quad x_i \quad \dots \quad x_n)$$

$$Y: \quad (y_1 \quad y_2 \quad \dots \quad y_i \quad \dots \quad y_n)$$

$$Z = X + Y$$

$$Z: \quad ((x_1 + y_1); (x_2 + y_2); \quad \dots \quad ; (x_i + y_i); \quad \dots \quad ; (x_n + y_n))$$

$$\bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$$

$$\bar{z} = \frac{\sum_{i=1}^{n} (x_i + y_i)}{n} = \frac{\sum_{i=1}^{n} x_i}{n} + \frac{\sum_{i=1}^{n} y_i}{n} = \bar{x} + \bar{y}$$

$$\bar{z} = \bar{x} + \bar{y}$$

- în cazul seriilor ponderate:

$$X: \quad \begin{pmatrix} x_1, & x_2, & \dots, & x_i, & \dots, & x_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

$$Y: \quad \begin{pmatrix} y_1, & y_2, & \dots, & y_i, & \dots, & y_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

$$Z = X + Y$$

$$Z: \quad \begin{pmatrix} x_1 + y_1, & x_2 + y_2, & \dots, & x_i + y_i, & \dots, & x_k + y_k \\ n_1, & n_2, & \dots, & n_i, & \dots, & n_k \end{pmatrix}$$

Scanned with OKEN Scanner

$$\bar{z} = \frac{\sum_{i=1}^{k} z_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} (x_i + y_i) n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} + \frac{\sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i} = \bar{x} + \bar{y}$$

$$\bar{z} = \bar{x} + \bar{y}$$

Dacă în suma a două variabile X şi Y prima se înmulţeşte cu o constantă a şi a doua cu b (a şi b putând fi pozitive sau negative), atunci media variabilei sumă $Z = aX + bY$ va fi:

$$\bar{z} = a\bar{x} + b\bar{y}$$

Astfel:

- în cazul seriilor simple:

$$\bar{z} = \frac{\sum_{i=1}^{n} z_i}{n} = \frac{\sum_{i=1}^{n} (ax_i + by_i)}{n} = \frac{\sum_{i=1}^{n} ax_i}{n} + \frac{\sum_{i=1}^{n} by_i}{n} = a\bar{x} + b\bar{y}$$

$$\bar{z} = a\bar{x} + b\bar{y}$$

- în cazul seriilor cu frecvenţă:

$$\bar{z} = \frac{\sum_{i=1}^{k} z_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} (ax_i + by_i) n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} ax_i n_i}{\sum_{i=1}^{k} n_i} + \frac{\sum_{i=1}^{k} by_i n_i}{\sum_{i=1}^{k} n_i} = a\bar{x} + b\bar{y}$$

deci $$\bar{z} = a\bar{x} + b\bar{y}$$

Dacă în relaţia de mai sus se substituie $a = 1$ şi $b = -1$ atunci se obţine:

$$Z = X - Y$$

şi $$\bar{z} = \bar{x} - \bar{y}$$

Scanned with OKEN Scanner

Această proprietate se utilizează în practică în problemele economice în care se cere determinarea nivelului mediu al unor fenomene complexe între componentele cărora există relaţii de însumare. De exemplu, volumul mediu al vânzărilor de mărfuri pe un locuitor se poate determina pe baza relaţiei (6.21), utilizând volumul mediu al vânzărilor mărfurilor nealimentare, al celor alimentare şi al alimentaţiei publice pe un locuitor, deoarece volumul total al desfacerilor de mărfuri este egal cu suma desfacerilor de mărfuri nealimentare, al celor alimentare şi al alimentaţiei publice.

12. *Media aritmetică a produsului a k variabile independente este egală cu produsul mediilor celor k variabile:*

$$\overline{X_1 \cdot X_2 \cdot \dots \cdot X_i \cdot \dots \cdot X_k} = \overline{X}_1 \cdot \overline{X}_2 \cdots \overline{X}_i \cdots \overline{X}_k \qquad (6.22)$$

Această proprietate se aplică în practică pentru determinarea nivelului mediu al unor fenomene complexe formate din fenomene simple între care şi fenomenul complex există o relaţie sub formă de produs. De exemplu, fondul de salarii este egal cu produsul dintre salariul mediu şi numărul de persoane angajate, volumul producţiei este egal cu produsul dintre productivitatea muncii şi numărul de muncitori.

Simplificarea calculelor ce se obţine prin aplicarea proprietăţilor mediei aritmetice o vom ilustra în continuare, printr-un exemplu concret.

În tabelul 6.4, se prezintă în mod sistematizat calculele intermediare necesare determinării mediei aritmetice, prin aplicarea uneia sau a mai multor proprietăţi ale ei.

Din calculele efectuate anterior a rezultat că $\bar{x} = 3013$ piese pe schimb.

Se vede din serie că $x_{min} = 2554$ şi $x_{max} = 345{,}5$ deci se verifică prima proprietate:

$$255 < 301 < 345$$

Proprietatea 3 se verifică imediat pe baza relaţiei (6.8):

$$\sum_{i=1}^{9} \bar{x} n_i = 301 \cdot 200 = 60\,200$$

$$\sum_{i=1}^{9} x_i n_i = 60\,200$$

Scanned with OKEN Scanner

Tabelul 6.4. Calculul simplificat al mediei aritmetice

Productivitatea muncii (piese/zi)	Numărul muncitorilor (n_i)	Mijloacele intervalelor (x_i)	$x_i n_i$	$(x_i - a)n_i$	$\frac{x_i - a}{k} n_i$	$\frac{n_i}{m}$	$(x_i - a)\frac{n_i}{m}$	$\frac{x_i - a}{k}\frac{n_i}{m}$	$\frac{x_i}{k} n_i$	$\frac{x_i}{k}\frac{n_i}{m}$	$x_i \frac{n_i}{m}$
0	1	2	3	4	5	6	7	8	9	10	11
255 - 265	5	260	1300	-200	-20	1	-40	-4	130	26	260
265 - 275	10	270	2700	-300	-30	2	-60	-6	270	54	540
275 - 285	20	280	5600	-400	-40	4	-80	-8	560	112	1120
285 - 295	40	290	11600	-400	-40	8	-80	-8	1160	232	2320
295 - 305	50	**300**	15000	0	0	10	0	0	1500	300	3000
305 - 315	30	310	9300	300	30	6	60	6	930	186	1860
315 - 325	25	320	8000	500	50	5	100	10	800	160	1600
325 - 335	10	330	3300	300	30	2	60	6	330	66	660
335 - 345	10	340	3400	400	40	2	80	8	340	68	680
Total	200	-	60200	200	20	40	40	4	6020	1204	12040

Date convenţionale

$$a = 300 \qquad k = 10 \qquad m = 5$$

Scanned with OKEN Scanner

Utilizând formula de calcul simplificat (6.15), prin aplicarea proprietăţii 6 se obţine:

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i}{k} n_i}{\sum_{i=1}^{s} n_i} \cdot k = \frac{6020}{200} \cdot 10 = 301 \quad piese/schimb$$

Prin aplicarea proprietăţilor 5 şi 6, rezultă:

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i - a}{k} n_i}{\sum_{i=1}^{s} n_i} \cdot k + a = \frac{20}{200} \cdot 10 + 300 = 301 \quad piese/schimb$$

Prin aplicarea proprietăţii 7, se obţine:

$$\bar{x} = \frac{\sum_{i=1}^{s} x_i \frac{n_i}{m}}{\sum_{i=1}^{s} \frac{n_i}{m}} = \frac{12040}{40} = 301 \quad piese/schimb$$

O altă simplificare a calculelor rezultă din aplicarea proprietăţilor 6 şi 7, astfel:

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i}{k} \frac{n_i}{m}}{\sum_{i=1}^{s} \frac{n_i}{m}} \cdot k = \frac{12040}{40} = 301 \quad piese/schimb$$

Cea mai mare simplificare a calculelor este obţinută prin aplicarea formulei (6.16):

Scanned with OKEN Scanner

$$\bar{x} = \frac{\sum_{i=1}^{s} \frac{x_i - a}{k} \cdot \frac{n_i}{m}}{\sum_{i=1}^{s} \frac{n_i}{m}} \cdot k + a = \frac{4}{40} \cdot 10 + 300 = 301 \quad piese/schimb$$

În cazul seriei cu frecvenţă relativă se pot efectua, de asemenea, simplificări ale calculului mediei aritmetice.

6.2.1.2. Avantajele şi dezavantajele mediei aritmetice

1. Media aritmetică este strict determinată printr-o formulă matematică precisă şi este uşor de calculat. Din punct de vedere teoretic, orice medie aritmetică este ponderată. În cazul seriilor statistice simple, fiecare variantă a variabilei este ponderată cu o pondere constantă şi egală cu 1. Deci media aritmetică simplă reprezintă un caz particular al mediei aritmetice ponderate.

2. Uneori, în practica statistică, nu se cunosc nivelurile individuale ale variabilei cercetate, ci se cunoaşte doar suma acestora şi numărul variantelor. În astfel de cazuri, media aritmetică, este singura medie ce se poate calcula.

3. Media aritmetică prezintă cea mai mică fluctuaţie de selecţie faţă de toate celelalte mărimi medii.

4. Fiind determinată pe baza unei formule precise, media aritmetică se pretează la calcule algebrice ulterioare. Astfel, dacă se cunosc mediile grupelor din cadrul unei colectivităţi şi volumul acestor grupe, se poate determina pe baza acestor informaţii media generală a colectivităţii.

Pe lângă aceste importante avantaje, media aritmetică prezintă şi unele *dezavantaje:*

1. Media aritmetică este foarte sensibilă la valorile extrem de mari şi la cele extrem de mici ale caracteristicii, iar în cazul seriilor care prezintă astfel de valori, ea este nereprezentativă.

2. Media aritmetică este inaplicabilă în cazul distribuţiilor cu asimetrie pronunţată. În aceste cazuri mediana sau modul dau rezultate mai semnificative.

3. Media aritmetică este inaplicabilă şi în cazul distribuţiilor în formă de U.

Scanned with OKEN Scanner

6.2.1.3. Media aritmetică a unei caracteristici alternative

Să considerăm o colectivitate statistică formată din n unităţi din care n_1 posedă nivelul x_1, iar n_2 posedă nivelul x_2 al caracteristicii.

Notând nivelul caracteristicii $x_1 = 1$, iar $x_2 = 0$, forma generală a unei distribuţii, în cazul unei caracteristici alternative, se prezintă în tabelul 6.5. unde

$$n_1 + n_2 = n \qquad p = \frac{n_1}{n} \qquad \text{iar} \qquad q = \frac{n_2}{n}$$

şi $$p + q = \frac{n_1}{n} + \frac{n_2}{n} = 1$$

Tabelul 6.5. Serie cu caracteristică alternativă

Variantele caracteristicii	Frecvenţa absolută n_i	Frecvenţa relativă f_i
$x_1 = 1$	n_1	p
$x_2 = 0$	n_2	q
Total	n	1

Media aritmetică a unei caracteristici alternative devine:

- în cazul utilizării frecvenţelor absolute:

$$\bar{x} = \frac{\sum_{i=1}^{2} x_i n_i}{\sum_{i=1}^{2} n_i} = \frac{1 \cdot n_1 + 0 \cdot n_2}{n_1 + n_2} = \frac{n_1}{n_1 + n_2} = \frac{n_1}{n} = p \qquad (6.23)$$

- în cazul utilizării frecvenţelor relative:

Scanned with OKEN Scanner

$$\bar{x} = \frac{\sum_{i=1}^{2} x_i f_i}{\sum_{i=1}^{2} f_i} = \frac{1 \cdot p + 0 \cdot q}{p+q} = \frac{p}{p+q} = p \qquad (6.24)$$

Prin urmare, media aritmetică a unei caracteristici alternative este egală cu frecvenţa relativă a unităţilor care posedă nivelul x_1 al caracteristicii.

6.2.2. Media armonică

Media armonică este o formă transformată a mediei aritmetice şi se determină ca o medie aritmetică inversă, calculată din mărimile inverse ale caracteristicii. Această medie a fost introdusă în statistică de către marele statistician şi economist *William Stanley Jevons* în anul 1874.

Dacă considerăm o serie cu frecvenţă:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_k \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_k \end{pmatrix}$$

atunci, conform definiţiei, media armonică notată cu $\bar{x}_h$ se va calcula astfel:

$$\bar{x}_h = \frac{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k}{\frac{1}{x_1}n_1 + \frac{1}{x_2}n_2 + \frac{1}{x_3}n_3 + \cdots + \frac{1}{x_i}n_i + \cdots + \frac{1}{x_k}n_k} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{1}{x_i} n_i}$$

deci

$$\bar{x}_h = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{1}{x_i} n_i} \qquad (6.25)$$

Relaţia (6.25) poartă denumirea de formula *mediei armonice ponderate.*

Scanned with OKEN Scanner

Ca şi în cazul celorlalte medii, atunci când frecvenţele tuturor nivelurilor individuale ale unei caracteristici sunt egale, acestea se pot neglija obţinându-se, din relaţia (6.25), formula *mediei armonice simple*. Substituind în (6.25) pe $n_i = c = constant$, rezultă:

$$\bar{x}_h = \frac{\sum_{i=1}^{k} c}{\sum_{i=1}^{k} \frac{c}{x_i}} = \frac{k \cdot c}{c \sum_{i=1}^{k} \frac{1}{x_i}} = \frac{k}{\sum_{i=1}^{k} \frac{1}{x_i}}$$

deci

$$\bar{x}_h = \frac{k}{\sum_{i=1}^{k} \frac{1}{x_i}} \qquad (6.26)$$

Din formulele (6.25) şi (6.26), se observă că media armonică se poate calcula numai atunci când toate variantele caracteristicii sunt valori pozitive ($x_i > 0$).

Media armonică este imposibil de calculat pentru o variabilă care are cel puţin una dintre variante egală cu zero, deoarece din punct de vedere matematic se ajunge la o nedeterminare.

În practică, dacă o serie este formată dintr-un număr redus de termeni, calculul mediei armonice se poate efectua pe baza unor formule echivalente şi simplificatoare:

- pentru serii formate din doi termeni:

$$\bar{x}_h = \frac{2}{\frac{1}{x_1} + \frac{1}{x_2}} = \frac{2}{\frac{x_1 + x_2}{x_1 x_2}} = \frac{2x_1 x_2}{x_1 + x_2}$$

- pentru serii formate din trei termeni:

$$\bar{x}_h = \frac{3}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3}} = \frac{3x_1 x_2 x_3}{x_1 x_2 + x_1 x_3 + x_2 x_3}$$

- pentru serii formate din patru termeni:

Scanned with OKEN Scanner

$$\bar{x}_h = \frac{4x_1x_2x_3x_4}{x_1x_2x_3 + x_1x_2x_4 + x_2x_3x_4 + x_1x_3x_4}$$

Din punct de vedere teoretic calculul mediei armonice cu ajutorul unor astfel de formule s-ar putea face pentru serii formate dintr-un număr oricât de mare de termeni. În practică, însă, aceste formule simplifică calculul numai pentru serii formate dintr-un număr redus de termeni, iar pe măsură ce acest număr creşte, calculele devin din ce în ce mai laborioase.

6.2.2.1. Proprietăţile mediei armonice

1. *Media armonică este o valoare internă a unei serii statistice.*

Pentru demonstrarea acestei proprietăţi considerăm cazul general al unei serii cu frecvenţă:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_k \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_k \end{pmatrix}$$

unde $x_1 < x_2 < x_3 < \cdots < x_i < \cdots < x_k$

iar mărimile inverse ale caracteristicii formează şirul descrescător:

$$\frac{1}{x_1} > \frac{1}{x_2} > \frac{1}{x_3} > \cdots > \frac{1}{x_i} > \cdots > \frac{1}{x_k}$$

Substituind în formula mediei armonice ponderate (6.25) variantele x_i cu $x_1 = x_{min}$, rezultă:

$$\bar{x}_h = \frac{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \cdots + \frac{n_i}{x_i} + \cdots + \frac{n_k}{x_k}} > \frac{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k}{\frac{n_1}{x_1} + \frac{n_2}{x_1} + \frac{n_3}{x_1} + \cdots + \frac{n_i}{x_1} + \cdots + \frac{n_k}{x_1}}$$

Această relaţie are loc deoarece:

Scanned with OKEN Scanner

$$\sum_{i=1}^{k} \frac{n_i}{x_i} < \sum_{i=1}^{k} \frac{n_i}{x_1}$$

întrucât $$\frac{n_1}{x_1} = \frac{n_1}{x_1}; \frac{n_2}{x_2} < \frac{n_2}{x_1}; \frac{n_3}{x_3} < \frac{n_3}{x_1}; \dots; \frac{n_k}{x_k} < \frac{n_k}{x_1}$$

deci $$\overline{x}_h > x_1$$

În mod analog, dacă substituim toate variantele x_i ale caracteristicii cu $x_k = x_{max}$, se obţine:

$$\overline{x}_h < x_k$$

rezultă: $$x_1 < \overline{x}_h < x_k \qquad (6.27)$$

2. *Media armonică este asociativă.*

Potrivit acestei proprietăţi, dacă într-o serie statistică se substituie un număr oarecare de termeni prin media lor armonică, atunci media armonică a seriei modificate este egală cu cea a seriei iniţiale.

Demonstrarea acestei proprietăţi o vom face tot pe baza cazului general al seriilor cu frecvenţă.

Considerăm seria:

$$X: \begin{pmatrix} x_1, x_2, x_3, \dots, x_k, \dots, x_s \\ n_1, n_2, n_3, \dots, n_k, \dots, n_s \end{pmatrix}$$

Primele k variante ale caracteristicii se înlocuiesc cu media lor armonică. Media armonică ponderată a seriei modificate devine:

$$\overline{x}_h = \frac{n_1 + n_2 + n_3 + \dots + n_i + \dots + n_k + n_{k+1} + \dots + n_s}{\frac{n_1}{\overline{x}_h} + \frac{n_2}{\overline{x}_h} + \frac{n_3}{\overline{x}_h} + \dots + \frac{n_k}{\overline{x}_h} + \frac{n_{k+1}}{x_{k+1}} + \dots + \frac{n_s}{x_s}} =$$

Scanned with OKEN Scanner

$$= \frac{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + \cdots + n_k + n_{k+1} + \cdots + n_s}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \cdots + \frac{n_k}{x_k}(n_1 + n_2 + n_3 + \cdots + n_k) + \frac{n_{k+1}}{x_{k+1}} + \cdots + \frac{n_s}{x_s}} =$$

$$= \frac{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k + n_{k+1} + \cdots + n_s}{\frac{n_1}{x_1} + \frac{n_2}{x_2} + \frac{n_3}{x_3} + \cdots + \frac{n_k}{x_k} + \frac{n_{k+1}}{x_{k+1}} + \cdots + \frac{n_s}{x_s}} = \frac{\sum_{i=1}^{s} n_i}{\sum_{i=1}^{s} \frac{n_i}{x_i}}$$

3. Din proprietatea 2, rezultă că *media armonică generală* $\bar{x}_h$ *a unei caracteristici dintr-o colectivitate împărţită în grupe este egală cu media armonică a mediilor armonice calculate pentru grupele ei* $\bar{x}_{h,i}$*:*

$$\bar{x}_h = \frac{\sum_{i=1}^{s} n_{i.}}{\sum_{i=1}^{s} \frac{1}{\bar{x}_{h,i}} n_{i.}} \qquad (6.28)$$

unde n_i reprezintă frecvenţa totală a grupei *i*.

4. *Dacă se micşorează frecvenţele unei serii de un anumit număr de ori* c, $c \neq 0$ *atunci media armonică a seriei modificate este egală cu media armonică a seriei iniţiale:*

$$\bar{x}_h = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}} = \frac{\sum_{i=1}^{k} \frac{n_i}{c}}{\sum_{i=1}^{k} \frac{n_i}{x_i}} = \frac{\frac{1}{c}\sum_{i=1}^{k} n_i}{\frac{1}{c}\sum_{i=1}^{k} \frac{n_i}{x_i}} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} \frac{n_i}{x_i}}$$

Menţionăm că *media armonică nu este translativă.*

Media armonică se aplică şi în determinarea dinamicii fenomenelor economice.

Scanned with OKEN Scanner

Pentru o mai bună înţelegere a esenţei şi rolului acestei medii în statistică, să considerăm câteva exemple.

Exemplul 1. Să presupunem că un produs agro-alimentar îşi modifică preţul în timpul unui trimestru de trei ori. Astfel, în prima lună cu 600 lei se puteau cumpăra 5 kg, în luna a doua cu aceeaşi sumă 4 kg, iar în luna a treia 3 kg. Se cere să se determine preţul mediu al acestui produs în trimestrul luat în calcul.

Dacă s-ar aplica media aritmetică a volumului fizic lunar am obţine:

$$\bar{x} = \frac{5+4+3}{3} = 4\ kg$$

$\bar{x} = 4\ kg$, ceea ce înseamnă că în medie într-o lună a trimestrului cu 600 lei se puteau cumpăra 4 kg din produsul respectiv, iar preţul mediu pe un kg ar fi fost egal cu 150 lei:

$$\frac{600}{4} = 150\ lei/kg$$

Preţul mediu, calculat după principiul mediei aritmetice, nu corespunde realităţii, deoarece nu ţine seama de cantitatea reală din produs vândută în fiecare lună. Dacă am cunoaşte aceste cantităţI, atunci preţul mediu s-ar putea calcula cu ajutorul mediei aritmetice ponderate.

De aceea, în acest caz calculul corect al preţului mediu se poate face cu ajutorul mediei armonice simple:

$$\bar{x}_h = \frac{n}{\sum_{i=1}^{3} \frac{1}{x_i}} = \frac{3}{\frac{1}{5}+\frac{1}{4}+\frac{1}{3}} = \frac{3}{\frac{12+15+20}{60}} = \frac{3 \cdot 60}{47} = \frac{180}{47} = 3{,}83\ kg$$

Deci, numărul mediu de kg ce se puteau cumpăra, în timpul trimestrului, cu 600 lei a fost de 3,83 kg, iar preţul mediu pe un kg, de aproximativ 156,65 lei.

Faptul că preţul dat de media armonică simplă este cel real se poate demonstra foarte simplu.

Să presupunem că în fiecare lună a trimestrului s-a vândut aceeaşi cantitate de produs Q, iar pe întregul trimestru s-a vândut cantitatea de $3Q$. Dacă notăm cu p preţul mediu, rezultă că pentru cantitatea $3Q$ kg s-a încasat valoarea $3pQ$ lei. În prima lună preţul unui kg de produs a fost egal cu 120 lei, ceea ce înseamnă că pentru toată cantitatea Q kg s-au încasat $120Q$ lei. În luna a doua, preţul unitar a

Scanned with OKEN Scanner

fost egal cu 150 lei şi suma încasată a fost de $150Q$ lei, în timp ce în luna a treia preţul mediu a fost de 200 lei, iar valoarea încasată de $200Q$ lei.

Determinând preţul mediu cu ajutorul mediei aritmetice ponderate, rezultă:

$$p = \frac{120Q + 150Q + 200Q}{3Q} = \frac{470Q}{3Q} = 156{,}66\ lei/kg$$

S-a obţinut deci acelaşi rezultat ca şi în cazul utilizării mediei armonice simple.

Exemplul 2. Să presupunem acum, pornind de la exemplul anterior, că s-au vândut în fiecare lună a trimestrului cantităţi diferite din produsul considerat şi anume: 4000 kg în prima lună; 3000 kg în luna a doua şi 1500 kg în luna a treia.

În acest caz, preţul mediu se poate determina cu ajutorul formulei mediei armonice ponderate, în care se utilizează drept pondere valoarea încasată din vânzarea produsului în fiecare lună a trimestrului.

$$\overline{x}_h = \frac{\sum_{i=1}^{3} n_i}{\sum_{i=1}^{3} \frac{1}{x_i} n_i} = \frac{120 \cdot 4000 + 150 \cdot 3000 + 200 \cdot 1500}{\frac{1}{120} \cdot 480000 + \frac{1}{150} \cdot 450000 + \frac{1}{120} \cdot 300000} = 144{,}58\ lei/kg$$

La acelaşi rezultat s-ar fi ajuns, dacă se determina preţul mediu pe baza mediei aritmetice a preţurilor din fiecare lună ponderate cu cantităţile corespunzătoare:

$$\overline{x} = \frac{\sum_{i=1}^{3} x_i n_i}{\sum_{i=1}^{3} n_i} = \frac{480000 + 450000 + 300000}{4000 + 3000 + 1500} = 144{,}58\ lei$$

6.2.2.2. Determinarea mediei armonice pe cale grafică

Pentru determinarea mediei armonice pe cale grafică, se ia un sistem de axe rectangulare în care pe abscisă se trec valorile caracteristicii x_i, iar pe ordonată, valorile inverse ale acesteia $y = \frac{1}{x}$. Se construieşte hiperbola $y_i = \frac{1}{x_i}$, în care $i = \overline{1,n}$.

Scanned with OKEN Scanner

Se determină media aritmetică (simplă sau ponderată după natura seriei luate în calcul) a mărimilor $y_i = \frac{1}{x_i}$.

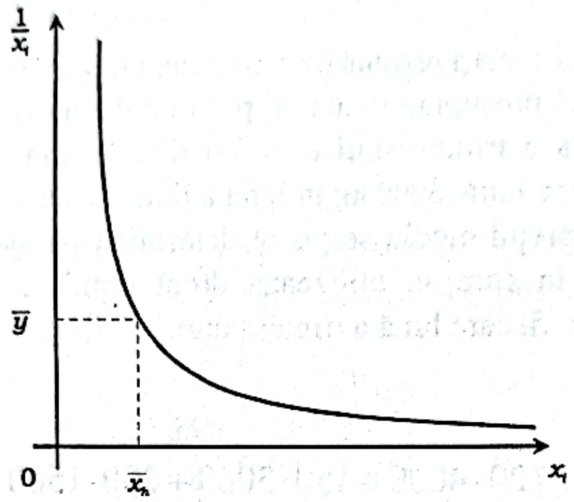


Figura 6.2. Determinarea mediei armonice

Din punctul de pe ordonată corespunzător mediei $\bar{y}$ se duce o paralelă la axa absciselor până întâlneşte hiperbola, iar din acest punct se coboară o perpendiculară pe abscisă. Abscisa piciorului perpendicularei este egală cu media armonică a variabilei cercetate (fig.6.2).

6.2.3. Media anarmonică

Numită şi *media contraarmonică*, *media anarmonică se calculează ca un raport între suma pătratelor nivelurilor individuale ale caracteristicii şi suma nivelurilor ei individuale.*

În cazul unei serii simple, se determină *media anarmonică simplă*, după formula:

Scanned with OKEN Scanner

$$\bar{x}_{ah} = \frac{\sum_{i=1}^{n} x_i^2}{\sum_{i=1}^{n} x_i} \tag{6.29}$$

Media anarmonică ponderată se utilizează în cazul seriilor cu frecvenţă şi se calculează ca un raport între suma pătratelor nivelurilor individuale ale caracteristicii ponderate cu frecvenţele lor şi suma nivelurilor individuale ale acesteia ponderate cu frecvenţele corespunzătoare:

$$\bar{x}_{ah} = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} x_i n_i} \tag{6.30}$$

Această medie se utilizează în practica statistică în probleme de prognoză.

6.2.4. Media geometrică

În cazul fenomenelor care se dezvoltă în progresie geometrică sau după o lege exponenţială, pentru determinarea nivelului lor mediu, se utilizează media geometrică, notată cu $\bar{x}_g$ şi introdusă în statistică de către *W.S. Jevons* în 1874.

Fie seria statistică cu frecvenţă:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_k \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_k \end{pmatrix}$$

Media geometrică a acestei serii este dată de relaţia:

$$\bar{x}_g = (x_1^{n_1} \cdot x_2^{n_2} \cdot x_3^{n_3} \cdots x_i^{n_i} \cdots x_k^{n_k})^{\frac{1}{\sum_{i=1}^{k} n_i}}$$

Scanned with OKEN Scanner

sau
$$\bar{x}_g = \sqrt[\sum_{i=1}^{k} n_i]{\prod_{i=1}^{k} x_i^{n_i}} \tag{6.31}$$

Formula (6.31) sau altele echivalente, reprezintă *media geometrică ponderată*. Se observă că media geometrică ponderată este egală cu rădăcina de ordinul $\sum_{i=1}^{k} n_i$ din produsul variantelor caracteristicii ridicate la o putere egală cu frecvenţa corespunzătoare.

Formula (6.31) arată că, dacă cel puţin o variantă a unei variabile este egală cu zero, atunci media geometrică a seriei este zero. De asemenea, dacă cel puţin una dintre variantele variabilei este mai mică decât zero, atunci media geometrică este un număr imaginar.

Logaritmând relaţia (6.31) rezultă:

$$log\,\bar{x}_g = \frac{1}{\sum_{i=1}^{k} n_i}(n_1 lg\,x_1 + n_2 lg\,x_2 + n_3 lg\,x_3 + \cdots + n_i lg\,x_i + \cdots + n_k lg\,x_k)$$

sau
$$lg\,\bar{x}_g = \frac{\sum_{i=1}^{k} n_i lg\,x_i}{\sum_{i=1}^{k} n_i} \tag{6.32}$$

Aceasta înseamnă că logaritmul mediei geometrice este egal cu media aritmetică a logaritmilor termenilor seriei. De aceea, media geometrică mai poartă şi denumirea de *medie logaritmică*.

Dacă $n_1 = n_2 = \cdots = n_i = \cdots = n_k = const$, atunci media geometrică a seriei devine:

$$\bar{x}_g = (x_1 \cdot x_2 \cdot x_3 \cdots x_i \cdots x_k)^{\frac{1}{k}} = \sqrt[k]{\prod_{1}^{k} x_i}$$

Scanned with OKEN Scanner

sau prin logaritmare: $$lg\,\bar{x}_g = \frac{\sum_{i=1}^{k} lg\,x_i}{k} \qquad (6.33)$$

În cazul unei serii formate numai din două valori x_1 şi x_2 media geometrică devine:

$$\bar{x}_g = \sqrt{x_1 x_2} \qquad (6.34)$$

Această formulă mai poartă şi denumirea de *medie proporţională*, deoarece rezultă din proporţia:

$$\frac{\bar{x}_g}{x_1} = \frac{x_2}{\bar{x}_g}$$

6.2.4.1. Proprietăţile mediei geometrice

1. *Media geometrică este o valoare internă a unei serii statistice.*
Pentru demonstrarea acestei proprietăţi să considerăm seria:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_s \end{pmatrix}$$

în care $$x_1 < x_2 < x_3 < \cdots < x_i < \cdots < x_s$$

Substituind în formula mediei geometrice ponderate variantele caracteristicii x_i cu x_1, se obţine: $\bar{x}_g > x_1$

Analog, substituind pe x_i cu x_s rezultă: $\bar{x}_g < x_s$

deci $$x_1 < \bar{x}_g < x_s$$

2. *Media geometrică este asociativă.*

Scanned with OKEN Scanner

Această proprietate exprimă faptul că, *dacă se substituie un anumit număr de termeni cu media lor geometrică şi apoi se calculează media geometrică a seriei modificate, aceasta coincide cu media geometrică a seriei iniţiale.*

Considerând seria:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_k, x_{k+1}, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_k, n_{k+1}, \ldots, n_s \end{pmatrix}$$

substituind primii k termeni cu media lor geometrică $\bar{x}_g$ în cadrul seriei iniţiale rezultă:

$$\bar{x}_g = \sqrt[\sum_1^k n_i]{\prod_{i=1}^{k} x_i^{n_i}} = \left(\prod_1^k x_i^{n_i}\right)^{\frac{1}{\sum_i^k n_i}}$$

iar

$$\sqrt[\sum_1^s n_i]{\bar{x}_g^{(n_1+n_2+\cdots+n_k)} x_{k+1}^{n_{k+1}} \cdots x_s^{n_s}} = \sqrt[\sum_1^s n_i]{\left(\prod_1^k x_i^{n_i}\right)^{\frac{n_1+n_2+\cdots+n_k}{\sum_1^k n_i}} \cdot x_{k+1}^{n_{k+1}} \cdots x_s^{n_s}} =$$

$$= \sqrt[\sum_1^s n_i]{x_1^{n_1} \cdot x_2^{n_2} \cdots x_k^{n_k} \cdots x_s^{n_s}} = \bar{x}_g$$

3. *Suma abaterilor logaritmilor variantelor unei variabile faţă de logaritmul mediei lor geometrice este egală cu zero:*

- pentru o serie simplă: $\bar{x}_g = \sqrt[n]{\prod_{i=1}^{n} x_i}$ sau $(\bar{x}_g)^n = \prod_{i=1}^{n} x_i$

Împărţind ambii membrii ai acestei relaţii prin $\bar{x}_g^n$ rezultă:

$$\frac{(\bar{x}_g)^n}{(\bar{x}_g)^n} = \frac{x_1}{\bar{x}_g} \cdot \frac{x_2}{\bar{x}_g} \cdot \frac{x_3}{\bar{x}_g} \cdots \frac{x_i}{\bar{x}_g} \cdots \frac{x_n}{\bar{x}_g} = 1$$

Scanned with OKEN Scanner

de unde, prin logaritmare se obţine:

$$(\ln x_1 - \ln \bar{x}_g) + (\ln x_2 - \ln \bar{x}_g) + \cdots + (\ln x_i - \ln \bar{x}_g) + \cdots + (\ln x_n - \ln \bar{x}_g) = 0$$

rezultă $$\sum_{i=1}^{n} (\ln x_i - \ln \bar{x}_g) = 0$$

- pentru o serie cu frecvenţă:

$$\bar{x}_g = \sqrt[\sum_{i=1}^{n} n_i]{\prod_{i=1}^{k} x_i^{n_i}} \qquad \text{iar} \qquad (\bar{x}_g)^{\sum_{i=1}^{k} n_i} = \prod_{i=1}^{k} x_i^{n_i}$$

Împărţind ambii membri ai acestei relaţii prin $\bar{x}_g^n$ rezultă:

$$\frac{x_1^{n_1}}{\bar{x}_g^{n_1}} \cdot \frac{x_2^{n_2}}{\bar{x}_g^{n_2}} \cdots \frac{x_i^{n_i}}{\bar{x}_g^{n_i}} \cdots \frac{x_k^{n_k}}{\bar{x}_g^{n_k}} = 1$$

Logaritmând expresia, rezultă:

$$(\ln x_1 - \ln \bar{x}_g) n_1 + (\ln x_2 - \ln \bar{x}_g) n_2 + \cdots + (\ln x_i - \ln \bar{x}_g) n_i + \cdots + (\ln x_k - \ln \bar{x}_g) n_k = 0$$

sau $$\sum_{i=1}^{k} (\ln x_i^{n_i} - \ln \bar{x}_g^{n_i}) n_i = 0$$

4. *Media geometrică a n valori pozitive, distincte este totdeauna mai mică decât media aritmetică calculată pe baza aceloraşi date.*

Diferenţa $\bar{x} - \bar{x}_g$ se măreşte pe măsură ce creşte diferenţa dintre valorile extreme ale variabilei, deoarece acestea influenţează în mai mare măsură media aritmetică decât media geometrică.

În cazul a două valori se obţine:

$$\bar{x} = \frac{x_1 + x_2}{2} \qquad \text{şi} \qquad \bar{x}_g = \sqrt{x_1 x_2}$$

Scanned with OKEN Scanner

$$\bar{x} - \bar{x}_g = \frac{x_1 + x_2}{2} - \sqrt{x_1 x_2} = \frac{1}{2}(\sqrt{x_1} - \sqrt{x_2})^2$$

deci $$\bar{x} - \bar{x}_g \geq 0$$

de unde rezultă:

$$\begin{cases} \bar{x} - \bar{x}_g > 0 & dacă\ x_1 \neq x_2 \\ \bar{x} - \bar{x}_g = 0 & dacă\ x_1 = x_2 \end{cases}$$

În cazul a trei valori diferenţa $\bar{x} - \bar{x}_g$ este egală cu:

$$\bar{x} - \bar{x}_g = \frac{x_1 + x_2 + x_3}{3} - \sqrt[3]{x_1 x_2 x_3} =$$

$$= \frac{1}{6}(\sqrt[3]{x_1} + \sqrt[3]{x_2} + \sqrt[3]{x_3})[(\sqrt[3]{x_1} - \sqrt[3]{x_2})^2 + (\sqrt[3]{x_2} - \sqrt[3]{x_3})^2 + (\sqrt[3]{x_3} - \sqrt[3]{x_1})^2] \geq 0$$

5. *Media geometrică a unui şir de valori pozitive şi distincte este întotdeuna mai mare decât media armonică.*

Fie un şir de valori pozitive şi distincte ale variabilei X:

$$x_1 < x_2 < x_3 < \cdots < x_i < \cdots < x_n$$

Pentru valorile extreme ale seriei are loc relaţia:

$$(x_1 - x_n)^2 > 0$$

Adunând la această inegalitate $4x_1 x_n$ rezultă:

$$x_1^2 + 2x_1 x_n + x_n^2 > 4x_1 x_n$$

Extrăgând apoi rădăcina pătrată se obţine:

$$x_1 + x_n > 2\sqrt{x_1 x_n} \qquad \text{şi} \qquad (x_1 + x_n)\sqrt{x_1 x_n} > 2x_1 x_n$$

Scanned with OKEN Scanner

de unde $$\sqrt{x_1 x_n} > \frac{2x_1 x_n}{x_1 + x_n}$$

dar $$\frac{2x_1 x_n}{x_1 + x_n} = \frac{2}{\frac{x_1 + x_n}{x_1 x_n}} = \frac{2}{\frac{1}{x_1} + \frac{1}{x_n}} = \bar{x}_h$$

iar $$\sqrt{x_1 x_n} = \bar{x}_g$$

Pentru cele două valori extreme ale variabilei X rezultă $\bar{x}_g > \bar{x}_h$. Procedând în mod similar pentru celelalte perechi x_2 şi x_{n-1}; x_3 şi x_{n-2} şi aşa mai departe, va rezulta de fiecare dată că media geometrică este mai mare decât cea armonică. Cu cât diferenţa dintre x_1 şi x_n este mai mare, cu atât diferenţa dintre media geometrică şi media armonică este mai mare şi invers.

6. *Puterea n a mediei geometrice calculată din k valori pozitive este egală cu media geometrică a puterii n a celor k valori.*

$$(\bar{x}_g)^n = \left(\sqrt[k]{x_1 x_2 \cdots x_k}\right)^n = \left(\sqrt[k]{\prod_{i=1}^{k} x_i}\right)^n$$

$$(\bar{x}_g)^n = \left[\left(\prod_{i=1}^{k} x_i\right)^{\frac{1}{k}}\right]^n \quad \text{sau} \quad (\bar{x}_g)^n = \left(\prod_{i=1}^{k} x_i^n\right)^{\frac{1}{k}}$$

$$(\bar{x}_g)^n = \sqrt[k]{x_1^n \cdot x_2^n \cdots x_k^n}$$

7. *Media geometrică a rapoartelor valorilor corespunzătoare a două serii formate din acelaşi număr de termeni este egală cu raportul dintre mediile geometrice ale celor două serii.*

Fie seriile: $X: \ (x_1, x_2, x_3, \ldots, x_n)$

$Y: \ (y_1, y_2, y_3, \ldots, y_n)$

Scanned with OKEN Scanner

Media geometrică a rapoartelor dintre termenii corespunzători ai celor două serii este:

$$\sqrt[n]{\frac{x_1}{y_1}\cdot\frac{x_2}{y_2}\cdot\frac{x_3}{y_3}\cdots\frac{x_n}{y_n}}=\sqrt[n]{\frac{\prod_{i=1}^{n}x_i}{\prod_{i=1}^{n}y_i}}=\frac{\bar{x}_g}{\bar{y}_g}$$

Raportul a două medii geometrice este egal cu media geometrică a rapoartelor variantelor celor două serii.

Fie două serii statistice:

$$X:\begin{pmatrix}x_1, & x_2,\ldots,x_i,\ldots,x_k\\ n_1, & n_2,\ldots,n_i,\ldots,n_k\end{pmatrix}$$

şi

$$Y:\begin{pmatrix}y_1, & y_2,\ldots,y_i,\ldots,y_k\\ n_1, & n_2,\ldots,n_i,\ldots,n_k\end{pmatrix}$$

atunci

$$\bar{x}_g=\sqrt[\sum_{i=1}^{k}n_i]{\prod_{i=1}^{k}x_i^{n_i}} \qquad \text{şi} \qquad \bar{y}_g=\sqrt[\sum_{i=1}^{k}n_i]{\prod_{i=1}^{k}y_i^{n_i}}$$

iar raportul acestor două medii va fi:

$$\frac{\bar{x}_g}{\bar{y}_g}=\frac{\sqrt[\sum_{i=1}^{k}n_i]{\prod_{i=1}^{k}x_i^{n_i}}}{\sqrt[\sum_{i=1}^{k}n_i]{\prod_{1}^{k}y_i^{n_i}}}=\frac{\left(\prod_{i=1}^{k}x_i^{n_i}\right)^{\frac{1}{\sum_{i=1}^{k}n_i}}}{\left(\prod_{i=1}^{k}y_i^{n_i}\right)^{\frac{1}{\sum_{i=1}^{k}n_i}}}=\left(\prod_{i=1}^{k}\frac{x_i}{y_i}\right)^{\frac{1}{\sum_{i=1}^{k}n_i}}$$

Menţionăm că această proprietate nu se verifică pentru media aritmetică.
Considerând din nou cele două serii rezultă:

Scanned with OKEN Scanner

$$\frac{\bar{x}_g}{\bar{y}_g} = \frac{\dfrac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}}{\dfrac{\sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i}} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} y_i n_i} \neq \frac{\sum_{i=1}^{k} \frac{x_i}{y_i} n_i}{\sum_{i=1}^{k} n_i}$$

8. *Media geometrică a două numere* x_1 *şi* x_2 *este egală cu media geometrică calculată din media aritmetică şi media lor armonică.*

$$\bar{x}_g = \sqrt{\bar{x}\,\bar{x}_h}$$

în care $\bar{x}_g = \sqrt{x_1 x_2}$; $\bar{x} = \dfrac{x_1 + x_2}{2}$; $\bar{x}_h = \dfrac{2x_1 x_2}{x_1 + x_2}$

deci $$\sqrt{\bar{x} \cdot \bar{x}_h} = \sqrt{\frac{x_1 + x_2}{2} \cdot \frac{2x_1 x_2}{x_1 + x_2}} = \sqrt{x_1 x_2} = \bar{x}_g$$

9. *Dacă o colectivitate este formată din* k *părţi componente, fiecare cuprinzând* n_i *termeni* $(i = \overline{1,k})$*, atunci media geometrică a întregii colectivităţi se poate calcula pe baza celor* k *medii geometrice parţiale.*

Fie: $\bar{x}_g$ - media geometrică a *colectivităţii*;

$\bar{x}_{ig}$ - media geometrică a grupei i

iar $$n = n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k$$

Într-adevăr:

$$n \lg \bar{x}_g = n_1 \lg \bar{x}_{1g} + n_2 \lg \bar{x}_{2g} + n_3 \lg \bar{x}_{3g} + \cdots + n_k \lg \bar{x}_{kg}$$

Scanned with OKEN Scanner

sau $$n \lg \bar{x}_g = \sum_{i=1}^{k} n_i \lg \bar{x}_{ig}$$

iar $$\lg \bar{x}_g = \frac{\sum_{i=1}^{k} n_i \lg \bar{x}_{ig}}{n} \qquad (6.35)$$

6.2.5. Media pătratică

Media pătratică $\bar{x}_p$ este o formă transformată a mediei aritmetice care se calculează prin extragerea rădăcinii pătrate din media aritmetică a pătratelor variantelor unei variabile.

În cazul unei serii simple, se va utiliza *media pătratică simplă*, dată de formula:

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}} \qquad (6.36)$$

În cazul unei serii cu frecvenţă, se calculează *media pătratică ponderată* pe baza relaţiei:

$$\bar{x}_p = \sqrt{\frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i}} \qquad (6.37)$$

Media pătratică este o medie cu aplicaţii speciale şi se utilizează în statistică atunci când este necesar să se acorde o importanţă mai mare variantelor mai mari ale unei anumite variabile.

Scanned with OKEN Scanner

6.2.5.1 Proprietăţile mediei pătratice

1. *Media pătratică este o medie internă a unei serii:*

$$x_{min} < \overline{x}_p < x_{max}$$

2. *Media pătratică a unei serii de repartiţie este întotdeauna mai mare decât media aritmetică:*

$$\overline{x}_p > \overline{x}$$

3. *Media pătratică este influenţată în mai mare măsură de variantele variabilei care au o valoare mai ridicată şi invers, deoarece, prin ridicare la pătrat, valorile mai mari vor deţine o pondere mai mare în cadrul sumei pătratelor lor decât valorile respective simple în cadrul sumei acestora.*

6.2.6. Generalizarea noţiunii de medie

În general se admite că media unei serii statistice este o funcţie simetrică a tuturor termenilor seriei care verifică *condiţiile*:

1) *Dacă toţi termenii seriei sunt egali, media este egală cu valoarea lor comună;*

2) *Media unei variabile este o valoare internă a acesteia.*

Aceste condiţii sunt îndeplinite când media este definită prin relaţia:

$$f(m) = \frac{1}{n}[f(x_1) + f(x_2) + \cdots + f(x_i) + \cdots + f(x_n)]$$

unde $f(x)$ este o funcţie (care variază totdeauna în acelaşi sens) monotonă, iar m media calculată.

Fie o serie statistică cu frecvenţă:

$$X: \begin{pmatrix} x_1, x_2, x_3, \ldots, x_i, \ldots, x_s \\ n_1, n_2, n_3, \ldots, n_i, \ldots, n_s \end{pmatrix}$$

Scanned with OKEN Scanner

Prin definiţie *momentul de ordinul* k al distribuţiei este egal cu media aritmetică a puterilor de ordinul k a abaterilor $x_i - a$, unde a este o constantă aleasă arbitrar:

$$m_k = \frac{\sum_{i=1}^{s}(x_i - a)^k n_i}{\sum_{i=1}^{s} n_i} \qquad i = \overline{1,s}$$

În raport cu valoarea constantei a, se deosebesc:

- *momente iniţiale*, când $a = 0$;
- *momente centrate*, când $a = \bar{x}$;
- *momente obişnuite*, când $\bar{x} \neq a \neq 0$.

Din momentul iniţial de ordinul k rezultă *media de ordinul* k.

$$\bar{x}_k^k = \frac{\sum_{i=1}^{s} x_i^k n_i}{\sum_{i=1}^{s} n_i} = \frac{x_1^k n_1 + x_2^k n_2 + x_3^k n_3 + \cdots + x_i^k n_i + \cdots + x_s^k n_s}{n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_s}$$

Rezultă

$$\bar{x}_k = \sqrt[k]{\frac{\sum_{i=1}^{s} x_i^k n_i}{\sum_{i=1}^{s} n_i}} = \left(\frac{\sum_{i=1}^{s} x_i^k n_i}{\sum_{i=1}^{s} n_i}\right)^{\frac{1}{k}} \tag{6.38}$$

Relaţia (6.38) reprezintă o *generalizare a valorilor medii*, care constituie cazuri particulare ale mediei de ordinul k, în condiţiile folosirii aceluiaşi sistem de ponderare.

În cazul seriilor simple, dacă $n_i = constant$, relaţia (6.38) devine:

$$\bar{x}_k = \sqrt[k]{\frac{\sum_{i=1}^{s} x_i^k}{s}} = \left(\frac{\sum_{i=1}^{s} x_i^k}{s}\right)^{\frac{1}{k}} \tag{6.39}$$

Scanned with OKEN Scanner

1. Dacă în relaţiile (6.38) şi (6.39) se face substituţia $k=-1$ se obţine *media armonică*:

$$\bar{x}_{-1}=\left(\frac{\sum_{i=1}^{s}x_i^{-1}n_i}{\sum_{i=1}^{s}n_i}\right)^{-1}=\frac{\sum_{i=1}^{s}n_i}{\sum_{i=1}^{s}\frac{1}{x_i}n_i}=\bar{x}_h$$

şi respectiv $$\bar{x}_{-1}=\left(\frac{\sum_{i=1}^{s}x_i^{-1}}{s}\right)^{-1}=\frac{s}{\sum_{i=1}^{s}\frac{1}{x_i}}=\bar{x}_h$$

2, Dacă se substituie $k=0$ în relaţiile (6.38) şi (6.39) se obţine *media geometrică*:

$$\bar{x}_0=\left(\frac{\sum_{i=1}^{s}x_i^{0}n_i}{\sum_{i=1}^{s}n_i}\right)^{\frac{1}{0}}=1^{\infty}$$

Pentru înlăturarea nedeterminării se logaritmează expresia mediei de ordinul k şi substituim pe $k=0$:

$$ln\,\bar{x}_0=\frac{1}{0}ln\frac{x_1^0n_1+x_2^0n_2+x_3^0n_3+\cdots+x_i^0n_i+\cdots+x_s^0n_s}{n_1+n_2+n_3+\cdots+n_i+\cdots+n_s}=\frac{0}{0}$$

Aplicând regula lui *l'Hôspital* rezultă:

$$\lim_{k\to 0}ln(\bar{x}_k)=\lim_{k\to 0}\frac{n_1x_1^k\,ln\,x_1+n_2x_2^k\,ln\,x_2+\cdots+n_sx_s^k\,ln\,x_s}{n_1+n_2+n_3+\cdots+n_i+\cdots+n_s}$$

Scanned with OKEN Scanner

pentru $k = 0$ rezultă $x_i^k = 1$:

$$\ln \bar{x}_0 = \frac{n_1 \ln x_1 + n_2 \ln x_2 + \cdots + n_s \ln x_s}{n_1 + n_2 + \cdots + n_s} = \frac{\ln x_1^{n_1} + \ln x_2^{n_2} + \cdots + \ln x_s^{n_s}}{n_1 + n_2 + \cdots + n_s} = \frac{\ln(x_1^{n_1} x_2^{n_2} \cdots x_s^{n_s})}{n_1 + n_2 + \cdots + n_s}$$

de unde $$\bar{x}_0 = \sqrt[\sum_{i=1}^{s} n_i]{x_1^{n_1} x_2^{n_2} \cdots x_s^{n_s}}$$

În mod analog, urmând acelaşi raţionament, se obţine formula mediei geometrice simple.

3. Substituind $k = 1$ în relaţiile (6.35) şi (6.36) se obţine *media aritmetică*:

$$\bar{x}_1 = \left(\frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i} \right)^{\frac{1}{1}} = \frac{\sum_{i=1}^{s} x_i n_i}{\sum_{i=1}^{s} n_i}$$

şi

$$\bar{x}_1 = \left(\frac{\sum_{i=1}^{s} x_i}{s} \right)^{\frac{1}{1}} = \frac{\sum_{i=1}^{s} x_i}{s}$$

4. Făcând substituţia $k = 2$ în relaţiile (6.38) şi (6.39) se obţine:

- *media pătratică ponderată*:

$$\bar{x}_2 = \left(\frac{\sum_{i=1}^{s} x_i^2 n_i}{\sum_{i=1}^{s} n_i} \right)^{\frac{1}{2}} = \sqrt{\frac{\sum_{i=1}^{s} x_i^2 n_i}{\sum_{i=1}^{s} n_i}} = \bar{x}_p$$

- *media pătratică simplă*:

Scanned with OKEN Scanner

$$\bar{x}_2 = \left(\frac{\sum_{i=1}^{s} x_i^2}{s} \right)^{\frac{1}{2}} = \sqrt{\frac{\sum_{i=1}^{s} x_i^2}{s}} = \bar{x}_p$$

5. Dacă se face substituţia $k = 3$ în relaţiile (6.35) şi (6.36) se obţine media cubică x_c:

• *media cubică ponderată:*

$$\bar{x}_3 = \left(\frac{\sum_{i=1}^{s} x_i^3 n_i}{\sum_{i=1}^{s} n_i} \right)^{\frac{1}{3}} = \sqrt[3]{\frac{\sum_{i=1}^{s} x_i^3 n_i}{\sum_{i=1}^{s} n_i}} = \bar{x}_c$$

• *media cubică simplă:*

$$\bar{x}_3 = \left(\frac{\sum_{i=1}^{s} x_i^3}{s} \right)^{\frac{1}{3}} = \sqrt[3]{\frac{\sum_{i=1}^{s} x_i^3}{s}} = \bar{x}_c$$

Pentru $k = 4,5,6,\ldots$ se obţin *medii de ordin superior*.

Generalizarea noţiunii de medie ne permite să stabilim şi raportul de mărime dintre diferite medii. Astfel, pe baza relaţiei (6.38), se pot ordona mediile uzuale în raport cu valorile corespunzătoare ale lui k:

Rezultă $$\bar{x}_h \leq \bar{x}_g \leq \bar{x} \leq \bar{x}_p \leq \bar{x}_c$$

Egalitatea dintre medii are loc numai în cazul $x_i = const$.

Scanned with OKEN Scanner

k	-1	0	1	2	3	...
Media	armonică $\bar{x}_{-1}$	geometrică $\bar{x}_0$	aritmetică $\bar{x}_1$	pătratică $\bar{x}_2$	cubică $\bar{x}_3$	...

Media geometrică reduce influenţa valorilor mari ale variabilei şi o măreşte pe cea a valorilor mici, explicând astfel de ce într-o serie media geometrică este inferioară mediei aritmetice:

$$\bar{x}_g < \bar{x}$$

Relaţia dintre media aritmetică şi media geometrică a două numere x şi y se poate ilustra pe cale grafică (fig. 6.3).

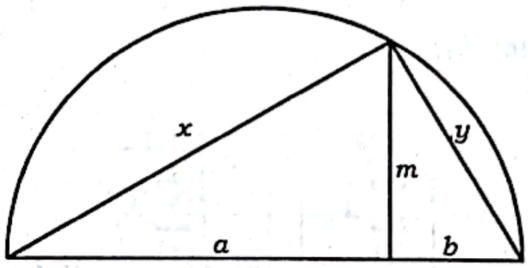


Figura 6.3. Relaţia dintre media aritmetică şi media geometrică a două numere x şi y

Diametrul semicercului este împărţit în două segmente a şi b de către perpendiculara m. Din geometria plană se ştie că $m = ab$. Dacă m este egală cu raza cercului, atunci $a = b = m$; şi acesta este cazul particular în care media aritmetică este egală cu media geometrică a celor două numere. Pentru cazurile când m cade în orice alt punct, diferit de centrul cercului, perpendiculara m va fi mai mică decât raza, iar media geometrică va fi mai mică decât media aritmetică.

Scanned with OKEN Scanner

6.3. Medii de poziţie

6.3.1. Modul

Modul este o mărime *medie de poziţie* care se poate determina numai în cazul seriilor cu frecvenţă. El se defineşte ca fiind *valoarea caracteristicii care se întâlneşte cel mai des în cadrul unei serii statistice, adică valoarea caracteristicii cu frecvenţă maximă*. De aceea modul, notat *Mo*, mai poartă şi denumirea de *valoarea cea mai frecventă* sau *cea mai probabilă* a caracteristicii sau, mai simplu, *valoarea dominantă*. *De exemplu*: venitul cel mai des întâlnit în cadrul unei anumite colectivităţi a populaţiei, ora de vârf dintr-o zi în privinţa transportului urban de călători într-un oraş, preţul cel mai des întâlnit (la care s-a vândut cea mai mare cantitate de produse) într-o anumită piaţă pentru un produs, talia cea mai frecventă în cadrul populaţiei dintr-o zonă geografică, ora la care centralele telefonice sunt cel mai încărcate, ora la care curba oboselii studenţilor atinge punctul de maxim etc.

A rezolva aceste probleme înseamnă a determina valoarea modală în cazul fiecărei distribuţii în parte.

Valoarea opusă modului, întâlnită în cazul distribuţiilor în formă de *J* şi *U*, se numeşte *antimod* şi reprezintă valoarea cea mai rar întâlnită în cadrul unei serii statistice sau minimul densităţii de repartiţie.

Cunoaşterea valorilor modală şi antimodală a unei variabile reale are o importanţă deosebită pentru factorii de decizie din diferite domenii ale activităţii economico-sociale.

Modul poate fi considerat ca un indicator corespunzător al tendinţei centrale numai în cazul distribuţiilor care se apropie de cea normală, deoarece în acest caz se întâlneşte un singur punct de concentrare maximă a variantelor caracteristicii şi, simetric de ambele părţi ale acestei valori, gradul de concentrare (reprezentat aici prin frecvenţă) scade. Distribuţiile cu un singur punct de maxim poartă denumirea de *distribuţii unimodale*. În practică, se întâlnesc şi distribuţii statistice cu două sau mai multe puncte de maxim, numite *distribuţii multimodale*. Distribuţiile cu două valori modale se numesc *distribuţii bimodale*.

În practică, distribuţiile multimodale pot să apară în două situaţii mai frecvente, şi anume când informaţiile iniţiale se referă la colectivităţi neomogene sau în cazul colectivităţilor omogene, dacă cantitatea de informaţie este insuficientă. În primul caz, nici una dintre valorile modale nu poate caracteriza cu o precizie satisfăcătoare tendinţa centrală pe întreaga scară de variaţie a variabilei. S-ar putea ca, în astfel de cazuri, diferite valori modale să caracterizeze diferite tipuri calitative care se diferenţiază în cadrul colectivităţilor, iar evidenţierea acestor tipuri să aibă o importanţă deosebită pentru conducerea activităţii în domeniul respectiv. În cel de-al

Scanned with OKEN Scanner

doilea caz, se poate stabili o valoare reprezentativă a modului, procedând la gruparea nivelurilor individuale ale caracteristicii pe intervale de variaţie, când datele iniţiale au fost grupate pe variante ale variabilei sau regrupând pe intervale mai mari informaţiile iniţiale, dacă ele se prezintă deja grupate pe intervale. Gruparea pe intervale sau regruparea se vor efectua prin încercări succesive, până când se va ajunge la stabilirea unei singure grupe modale.

Determinarea valorii reale a modului pentru o distribuţie este dificilă deoarece modul este o valoare medie care variază chiar în cazul aceloraşi informaţii în funcţie de variaţia mărimii intervalului de grupare şi a limitelor acestui interval. Valoarea reală a modului se poate determina cu exactitate numai în cazul în care numărul variantelor unei variabile tinde spre infinit, iar mărimea intervalelor se micşorează din ce în ce mai mult. În practică însă, determinarea unor valori aproximative ale modului este satisfăcătoare.

Determinarea modului se face în mod diferit pentru variabilele continue faţă de variabilele discrete.

6.3.1.1. Determinarea modului în cazul variabilelor discrete

Pentru o serie care prezintă variaţia unei variabile discrete pe variante ale caracteristicii, modul este egal cu varianta care are frecvenţă maximă. În acest caz, modul este bine determinat şi este egal cu un nivel real al caracteristicii care rezultă imediat pe baza citirii tabelului în care se prezintă seria.

Tabelul 6.14. Distribuţia studenţilor după nota obţinută la examen

Nota	Numărul de studenţi	% faţă de total
1	-	-
2	2	1
3	6	3
4	12	6
5	20	10
6	24	12
7	**76**	**38**
8	40	20
9	12	6
10	8	4
Total	200	100

$Mo = 7$

Scanned with OKEN Scanner

Să considerăm pentru exemplificare distribuţia a 100 de studenţi după nota obţinută la examenul de *Statistică* (tabelul 6.14). Potrivit regulamentului de funcţionare a învăţământului superior, nivelul de pregătire a studenţilor se apreciază de către cadrele didactice cu nota de la 10 la 1, numai prin numere întregi. În acest caz, nota obţinută la examene este o variabilă discretă definită pe intervalul de la 1 la 10.

Citind notele din tabel, se observă că nota obţinută de către cei mai mulţi studenţi este 7. Această notă a fost obţinută de către 76 de studenţi, ceea ce reprezintă 38% din totalul studenţilor, în timp ce notele mai mici, ca şi cele mai mari au fost obţinute de către un număr de studenţi din ce în ce mai mic, pe măsură ce ne depărtăm de nota 7 în ambele direcţii.

Modul se poate determina şi pe cale grafică ca fiind egal cu abscisa punctului care are ordonata maximă în cadrul poligonului de frecvenţă.

6.3.1.2. Determinarea modului în cazul unei variabile continue

6.3.1.2.1. Determinarea modului în cazul seriilor cu intervale egale

Riguros vorbind, modul reprezintă valoarea caracteristicii căreia îi corespunde densitatea maximă a unei distribuţii. Pentru distribuţiile continue, informaţiile iniţiale se grupează pe intervale de variaţie. Procedând ca în cazul variabilelor discrete, în condiţiile variabilelor continue se identifică *intervalul modal*, adică acel interval care are frecvenţa maximă. Problema de bază, care trebuie să fie rezolvată acum, este determinarea nivelului modal al caracteristicii în cadrul acestui interval. Cea mai simplă modalitate de determinare aproximativă a valorii modale este considerarea că *modul este egal cu centrul intervalului modal* cu condiţia ca, în cadrul acestui interval, variabila să fie distribuită în mod uniform, liniar. Desigur, uniformitatea distribuţiei în cadrul intervalului modal nu poate fi argumentată nici teoretic, nici practic. Mai mult decât atât, această ipoteză este în contradicţie cu însăşi definiţia modului.

Pentru determinarea cu exactitate a modului, este necesar să se cunoască în prealabil densitatea de repartiţie.

Să presupunem că porţiunea curbei din vecinătatea modului unei variabile continue variază după o parabolă de gradul doi:

$$f(x) = a + bx + cx^2$$

Considerăm, de asemenea, că intervalele de variaţie ale caracteristicii sunt egale. Pentru intervalul modal şi intervalele vecine lui se cunosc limitele acestora şi frecvenţele relative:

Scanned with OKEN Scanner

$x_{k-1} - x_k$	f_{k-1}
$x_k - x_{k+1}$	f_k
$x_{k+1} - x_{k+2}$	f_{k+1}

unde

$$f_{k-1} < f_k > f_{k+1}$$

Deoarece modul corespunde maximului densităţii de repartiţie, pentru determinarea lui egalăm prima derivată a funcţiei cu zero:

$$f'(x) = b + 2cx = 0$$

de unde

$$x = -\frac{b}{2c} = Mo \qquad (6.40)$$

Grafic această densitate şi valoarea modului se prezintă în figura 6.4.

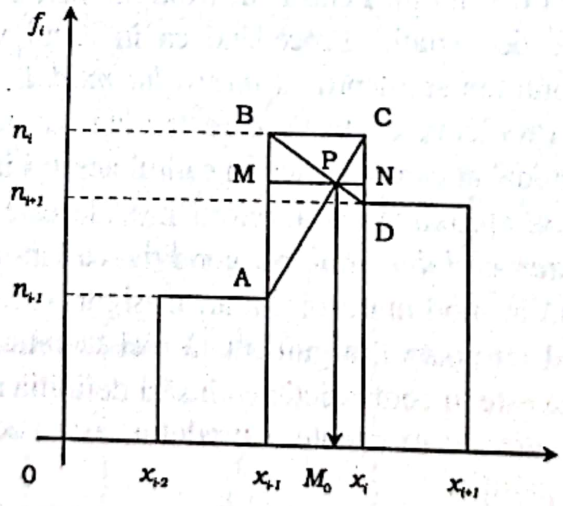


Figura 6.4. Densitatea repartiţiei şi valoarea modului

Pentru determinarea modului se va ajusta distribuţia teoretică în intervalul (x_{k-1}, x_{k+2}), cu ajutorul arcului de parabolă. Ajustarea distribuţiei se poate face aplicând metoda ariilor, astfel:

Scanned with OKEN Scanner

$$\begin{cases} \int_{x_{k-1}}^{x_k} f(x)dx = f_{k-1} \\ \int_{x_k}^{x_{k+1}} f(x)dx = f_k \\ \int_{x_{k+1}}^{x_{k+2}} f(x)dx = f_{k+1} \end{cases} \tag{6.41}$$

Pe baza sistemului (6.41) se pot determina parametrii a, b, şi c. Ţinând seama de faptul că cele trei intervale de variaţie a caracteristicii sunt egale, calculul parametrilor parabolei se poate simplifica făcând următoarea schimbare de variabilă:

$$y = \frac{x - x_k}{d} \tag{6.42}$$

unde $d = x_k - x_{k-1}$ este egal cu mărimea intervalului de grupare.

Substituind succesiv valorile lui x în relaţia (6.42) rezultă noile limite de integrare:

când $x = x_{k-1}$ $\quad y = \frac{x_{k-1} - x_k}{d} = \frac{-d}{d} = -1$

$x = x_k$ $\quad y = \frac{x_k - x_k}{d} = 0$

$x = x_{k+1}$ $\quad y = \frac{x_{k+1} - x_k}{d} = \frac{d}{d} = 1$

$x = x_{k+2}$ $\quad y = \frac{x_{k+2} - x_k}{d} = \frac{2d}{d} = 2$

iar

$$f_y = a + b(yd + x_k) + c(yd + x_k)^2 = a + bx_k + cx_k^2 + dy(b + 2cx_k)y + cd^2y^2$$

Scanned with OKEN Scanner

$$\begin{cases} \int_{x_{k-1}}^{x_k} f(x)dx = f_{k-1} \\ \int_{x_k}^{x_{k+1}} f(x)dx = f_k \\ \int_{x_{k+1}}^{x_{k+2}} f(x)dx = f_{k+1} \end{cases} \tag{6.41}$$

Pe baza sistemului (6.41) se pot determina parametrii a, b, şi c. Ţinând seama de faptul că cele trei intervale de variaţie a caracteristicii sunt egale, calculul parametrilor parabolei se poate simplifica făcând următoarea schimbare de variabilă:

$$y = \frac{x - x_k}{d} \tag{6.42}$$

unde $d = x_k - x_{k-1}$ este egal cu mărimea intervalului de grupare.

Substituind succesiv valorile lui x în relaţia (6.42) rezultă noile limite de integrare:

când $x = x_{k-1}$ $\quad y = \frac{x_{k-1} - x_k}{d} = \frac{-d}{d} = -1$

$x = x_k$ $\quad y = \frac{x_k - x_k}{d} = 0$

$x = x_{k+1}$ $\quad y = \frac{x_{k+1} - x_k}{d} = \frac{d}{d} = 1$

$x = x_{k+2}$ $\quad y = \frac{x_{k+2} - x_k}{d} = \frac{2d}{d} = 2$

iar

$$f_y = a + b(yd + x_k) + c(yd + x_k)^2 = a + bx_k + cx_k^2 + dy(b + 2cx_k)y + cd^2y^2$$

Scanned with OKEN Scanner

Sistemul (6.42) devine:

$$\begin{cases} \int_{-1}^{0} f(y)dy = (a+bx_k+cx_k^2) - \frac{d}{2}(b+2cx_k) + \frac{cd^2}{3} = f_{k-1} \\ \int_{0}^{1} f(y)dy = (a+bx_k+cx_k^2) + \frac{d}{2}(b+2cx_k) + \frac{cd^2}{3} = f_k \\ \int_{1}^{2} f(y)dy = (a+bx_k+cx_k^2) + \frac{3d}{2}(b+2cx_k) + \frac{7}{3}cd^2 = f_{k+1} \end{cases} \quad (6.43)$$

Scăzând apoi prima ecuaţie din a doua şi pe a doua din a treia se obţine:

$$\begin{cases} f_k - f_{k-1} = d(b+2cx_k) \\ f_{k+1} - f_k = d(b+2cx_k) + 2cd^2 \end{cases} \quad (6.44)$$

Scăzând apoi în (6.44) prima ecuaţie din cea de a doua rezultă:

$$2cd^2 = f_{k+1} - 2f_k + f_{k-1} \quad (6.45)$$

$$y = -\frac{f_k - f_{k-1}}{(f_{k+1} - f_k) - (f_k - f_{k-1})} = \frac{f_k - f_{k-1}}{(f_k - f_{k+1}) + (f_k - f_{k-1})}$$

Din relaţiile (6.40) şi (6.44) se obţine:

$$x = x_k + dy$$

deci $$Mo = x_k + d\frac{f_k - f_{k-1}}{(f_k - f_{k+1}) + (f_k - f_{k-1})}$$

Dacă notăm: $f_k - f_{k-1} = \Delta_1$ *si* $f_k - f_{k+1} = \Delta_2$ formula modului devine:

$$Mo = x_k + d\frac{\Delta_1}{\Delta_1 + \Delta_2} \quad (6.46)$$

Scanned with OKEN Scanner

unde: x_k = limita inferioară a intervalului modal;
d = mărimea intervalului în cazul seriilor cu intervale egale;
Δ_1 = diferenţa dintre frecvenţa intervalului modal şi a intervalului anterior;
Δ_2 = diferenţa dintre frecvenţa intervalului modal şi a intervalului următor.

Formula (6.46) se aplică în practică pentru determinarea *modului* în cazul seriilor cu intervale egale de variaţie a caracteristicii. Din această formulă rezultă că valoarea *modului* depinde numai de valorile caracteristicii din intervalul modal, de frecvenţa acestui interval şi a celor două vecine lui. Deoarece nu ţine seama de toate nivelurile individuale ale caracteristicii, *modul* este o medie aproximativă şi nu se pretează la calcule algebrice ulterioare.

Pe lângă aceste neajunsuri, *modul* prezintă avantajul că este o valoare cu semnificaţie concretă, uşor de calculat şi poate înlocui cu succes media aritmetică în cazul distribuţiilor simetrice cu intervale extreme deschise.

Valoarea *modului* pentru o anumită serie este aceeaşi, indiferent dacă se iau în calcul frecvenţele absolute sau cele relative.

Pentru ilustrarea calculului *modului* în cazul distribuţiilor cu intervale egale de variaţie a caracteristicii, considerăm datele din tabelul 6.15.

Citind coloana n_i sau f_i se observă că intervalul modal este (295 - 305), deoarece are frecvenţă maximă.

În cazul frecvenţelor absolute:

$$x_k = 295;\ d = 10;\ \Delta_1 = 50 - 40 = 10;\ \Delta_2 = 50 - 30 = 20$$

$$Mo = 295 + 10\frac{10}{10+20} = 298{,}33 \ \textit{piese/muncitor}$$

În cazul frecvenţelor relative:

$$x_k = 295;\ d = 10;\ \Delta_1 = 25{,}0 - 20{,}0 = 5{,}0;\ \Delta_2 = 25{,}0 - 15{,}0 = 10{,}0$$

$$Mo = 295 + 10\frac{5}{5+10} = 298{,}33 \ \textit{piese/muncitor}$$

Scanned with OKEN Scanner

Tabelul 6.15. Distribuţia muncitorilor după nivelul productivităţii muncii

Productivitatea muncii (piese/schimb)	Numărul muncitorilor n_i	f_i %	F%
255 - 265	5	2,5	2,5
265 - 275	10	5,0	7,5
275 - 285	20	10,0	17,5
285 - 295	40	20,0	37,5
295 - 305	50	25,0	62,5
305 - 315	30	15,0	77,5
315 - 325	25	12,5	90,0
325 - 335	10	5,0	95,0
335 - 345	10	5,0	100,0
Total	200	100,0	-

Mo = 298,33 piese pe un muncitor ne arată că productivitatea medie a muncii celor 200 de muncitori este egală cu 298,33 piese într-un schimb.

6.3.1.2.1.1. Determinarea grafică a modului

O metodă simplă de determinare a modului este interpolarea pe cale grafică.

Pentru aceasta se reprezintă grafic intervalul modal şi intervalele vecine prin coloane. Se uneşte, în diagonală, colţul din dreapta de sus al coloanei corespunzătoare intervalului anterior celui modal cu colţul din dreapta sus al coloanei ce reprezintă intervalul modal; apoi se uneşte colţul din stânga de sus al ultimei coloane cu colţul din stânga de sus al coloanei din mijloc. Abscisa punctului de intersecţie a celor două diagonale este egală cu modul (fig. 6.5).

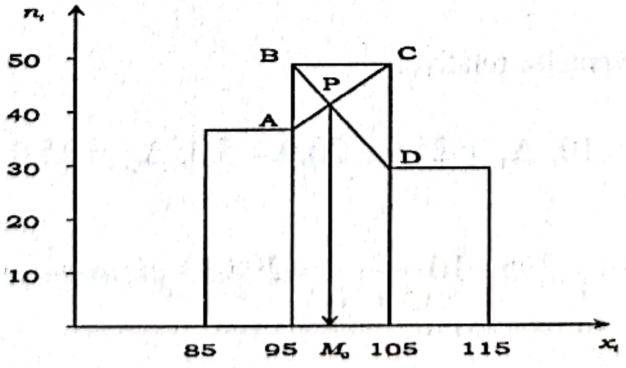
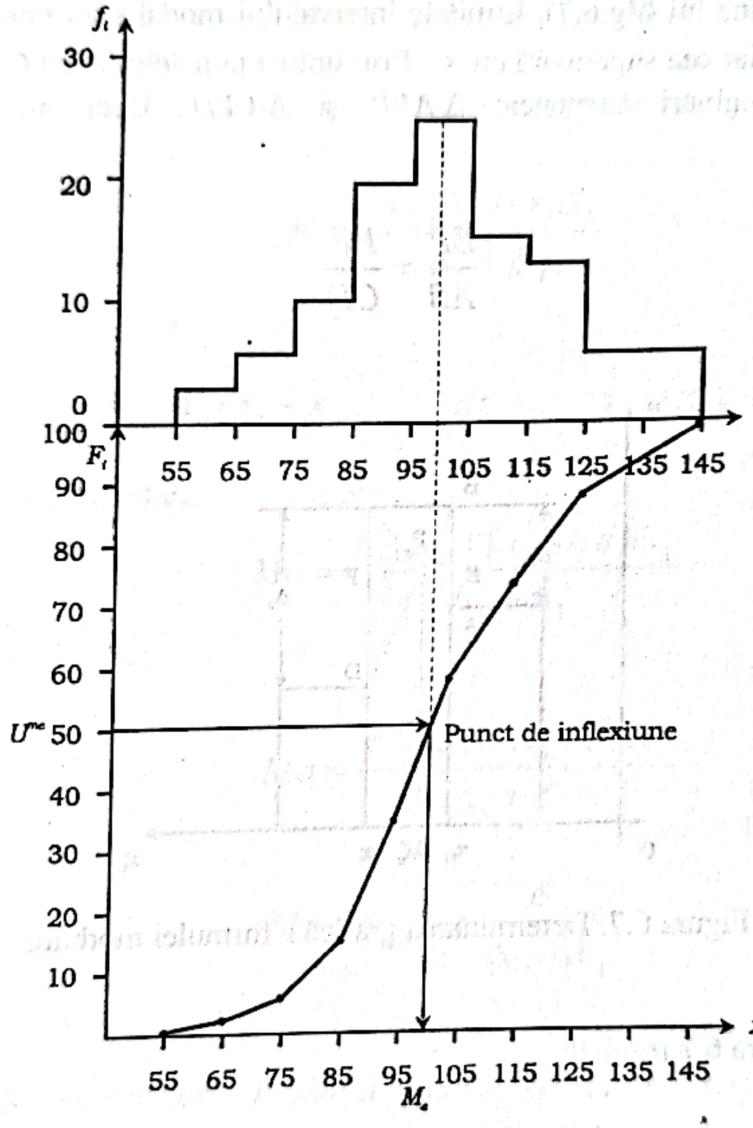


Figura 6.5. Determinarea modului prin interpolare grafică

Scanned with OKEN Scanner

O altă modalitate de determinare pe cale grafică a modului se obţine pornind de la curba frecvenţelor cumulate (fig. 6.6). În acest caz modul este egal cu abscisa punctului de inflexiune a curbei.



inv. 149701

Figura 6.6. Histograma, curba frecvenţelor cumulate şi modul

Scanned with OKEN Scanner

6.3.1.2.1.2. Stabilirea formulei de calcul al modului în cazul seriilor cu intervale egale

Să considerăm, din histogramă, coloanele corespunzătoare intervalului modal şi intervalelor vecine lui (fig.6.7). Limitele intervalului modal s-au notat astfel: limita inferioară cu x_{i-1} iar cea superioară cu x_i. Prin unirea punctelor A cu C şi B cu D s-au format două triunghiuri asemenea: $\Delta\, ABP$ şi $\Delta\, CPD$. Deci are loc relaţia de proporţionalitate:

$$\frac{EP}{AB} = \frac{PF}{CD}$$

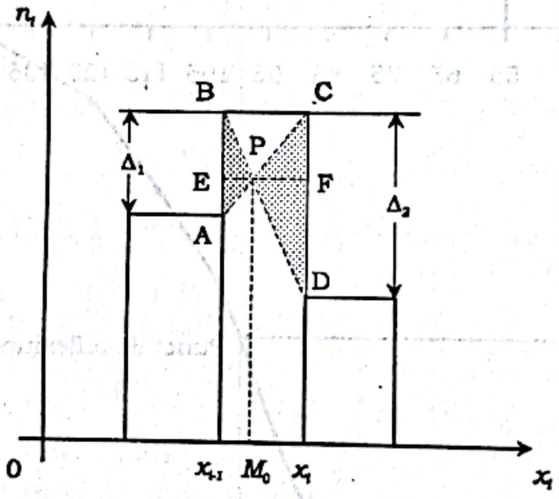


Figura 6.7. Determinarea grafică a formulei modului

Din figura 6.7 rezultă:

$$EP = Mo - x_{i-1} \qquad AB = \Delta_1$$

$$PF = x_i - Mo \qquad CD = \Delta_2$$

Ţinând seama de aceste relaţii se obţine:

Scanned with OKEN Scanner

$$\frac{Mo - x_{i-1}}{\Delta_1} = \frac{x_i - Mo}{\Delta_2}$$

deci $$\Delta_2(Mo - x_{i-1}) = \Delta_1(x_i - Mo)$$

$$Mo\Delta_2 - x_{i-1}\Delta_2 = x_i\Delta_1 - Mo\Delta_1$$

şi $$Mo = \frac{x_{i-1}\Delta_2 + x_i\Delta_1}{\Delta_1 + \Delta_2}$$

dar $d = x_i - x_{i-1}$ iar $x_i = x_{i-1} + d$

Modul va fi egal cu:

$$Mo = \frac{x_{i-1}\Delta_2 + (x_{i-1} + d)\Delta_1}{\Delta_1 + \Delta_2}$$

şi $$Mo = \frac{x_{i-1}(\Delta_1 + \Delta_2) + d\Delta_1}{\Delta_1 + \Delta_2}$$

$$Mo = x_{i-1} + d\frac{\Delta_1}{\Delta_1 + \Delta_2}$$

6.3.1.2.1.3. Determinarea modulului în cazul seriilor cu intervale neegale

În practică se întâlnesc de multe ori serii statistice cu intervale de variaţie neegale. Stabilirea formulei de calcul a modului şi în această situaţie se face pe baza aceloraşi principii ca şi în cazul seriilor cu intervale egale. Considerăm, din cadrul unei serii cu intervale neegale, intervalul modal şi intervalele vecine lui, notate astfel: (x_{k-1}, x_k); (x_k, x_{k+1}) şi (x_{k+1}, x_{k+2}) iar frecvenţele corespunzătoare cu n_{k-1}, n_k şi n_{k+1} sau f_{k-1}, f_k şi f_{k+1}.

Sistemul (6.41), care desemnează suprafeţele celor trei coloane mărginite în partea superioară de parabola $f(x)$, devine:

$$\begin{cases} \int_{x_{k-1}}^{x_k}(a+bx+cx^2)dx = f_{k-1} \\ \int_{x_k}^{x_{k+1}}(a+bx+cx^2)dx = f_k \\ \int_{x_{k+1}}^{x_{k+2}}(a+bx+cx^2)dx = f_{k+1} \end{cases} \quad (6.47)$$

Calculând integralele rezultă:

$$\begin{cases} f_{k-1} = a(x_k - x_{k-1}) + \frac{b}{2}(x_k^2 - x_{k-1}^2) + \frac{c}{3}(x_k^3 - x_{k-1}^3) \\ f_k = a(x_{k+1} - x_k) + \frac{b}{2}(x_{k+1}^2 - x_k^2) + \frac{c}{3}(x_{k+1}^3 - x_k^3) \\ f_{k+1} = a(x_{k+2} - x_{k+1}) + \frac{b}{2}(x_{k+2}^2 - x_{k+1}^2) + \frac{c}{3}(x_{k+2}^3 - x_{k+1}^3) \end{cases} \quad (6.48)$$

Rezolvând sistemul (6.48) se obţin valorile celor doi parametri b şi c de care depinde valoarea modului. Este recomandabil ca, înainte de rezolvarea acestui sistem, să se substituie variantele variabilei X cu nivelurile corespunzătoare din problema propusă spre rezolvare şi cu frecvenţele respective.

Cunoscând expresiile parametrilor b şi c le putem substitui în relaţia (6.40).

Formula (6.40) se aplică în practică pentru determinarea modului în cazul unei distribuţii pe intervale, indiferent dacă acestea sunt egale sau neegale. Deoarece în cazul distribuţiilor pe intervale neegale densităţile lor nu sunt egale cu frecvenţele relative, în stabilirea intervalului modal se vor lua în consideraţie densităţile distribuţiilor.

6.3.1.3. Proprietăţile modului

1. *Modul este cuprins între cea mai mică şi cea mai mare valoare a caracteristicii.*

$$x_{min} < Mo < x_{max}$$

2. *Valoarea modului unei serii nu se schimbă dacă frecvenţele (absolute sau cele relative) ale acesteia se împart la o anumită cantitate constantă m diferită de zero.* Astfel, în cazul unor serii cu intervale egale se obţine:

$$x_k + d\frac{\frac{n_k}{m} - \frac{n_{k-1}}{m}}{\frac{n_k}{m} - \frac{n_{k-1}}{m} + \frac{n_k}{m} - \frac{n_{k+1}}{m}} = x_k + d\frac{n_k - n_{k-1}}{(n_k - n_{k-1}) + (n_k - n_{k+1})} = Mo$$

3. *Dacă se măreşte sau se micşorează fiecare nivel individual al caracteristicii unei serii cu o anumită constantă a, atunci valoarea modului se măreşte sau se micşorează cu acea constantă:*

$$(x_k \pm a) + [(x_{k+1} \pm a) - (x_k \pm a)]\frac{\Delta_1}{\Delta_1 + \Delta_2} = x_k + d\frac{\Delta_1}{\Delta_1 + \Delta_2} \pm a = Mo \pm a$$

4. *Dacă se înmulţeşte sau se împarte fiecare nivel individual al caracteristicii unei serii printr-o constantă c diferită de zero, atunci modul se măreşte sau, respectiv, se micşorează de acelaşi număr de ori:*

$$\frac{x_k}{c} + \frac{x_{k+1} - x_k}{c} \cdot \frac{\Delta_1}{\Delta_1 + \Delta_2} = \frac{1}{c}(x_k + d\frac{\Delta_1}{\Delta_1 + \Delta_2}) = \frac{Mo}{c}$$

şi

$$cx_k + cd\frac{\Delta_1}{\Delta_1 + \Delta_2} = c(x_k + d\frac{\Delta_1}{\Delta_1 + \Delta_2}) = cMo$$

Folosind concomitent proprietăţile 2, 3 şi 4 rezultă o formulă de calcul simplificat al modului:

$$\frac{x_k - a}{c} + (\frac{x_{k+1}}{c} - \frac{a}{c})\frac{\frac{n_k}{m} - \frac{n_{k-1}}{m}}{\frac{n_k}{m} - \frac{n_{k-1}}{m} + \frac{n_k}{m} - \frac{n_{k+1}}{m}} = \frac{x_k - a}{c} + \frac{d}{c}\frac{\Delta_1}{\Delta_1 + \Delta_2} =$$

$$= \frac{1}{c}(x_k + d\frac{\Delta_1}{\Delta_1 + \Delta_2}) - \frac{a}{c} = \frac{Mo - a}{c}$$

Scanned with OKEN Scanner

6.3.2. Mediana

Mediana sau *valoarea mijlocie*, notată *Me*, *este* definită ca *valoarea caracteristicii, ordonate crescător sau descrescător a nivelurilor sale individuale, care are proprietatea ca numărul variantelor mai mari decât ea să fie egal cu numărul variantelor mai mici.*

Aceasta înseamnă că mediana unei serii statistice corespunde valorii echiprobabile a unei variabile X şi este definită de relaţia:

$$P(X < Me) = P(X > Me) = \frac{1}{2}$$

$$P(X > x) \geq \frac{1}{2} \quad pentru\ x > Me$$

$$P(X < x) \leq \frac{1}{2} \quad pentru\ x < Me$$

Din punct de vedere analitic mediana poate fi definită ca fiind egală cu varianta x_i a caracteristicii X, pentru care perpendiculara ridicată pe axa absciselor din punctul x_i împarte suprafaţa mărginită de curba frecvenţelor în două părţi egale (fig. 6.8).

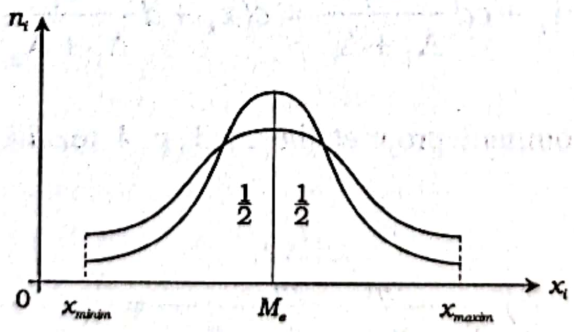


Figura 6.8. Curbe ale frecvenţelor şi locul medianei

Din punct de vedere teoretic, calculul medianei este simplu şi nu ridică dificultăţi deosebite.

Scanned with OKEN Scanner

Din punct de vedere practic în determinarea medianei se disting mai multe cazuri în funcţie de modul de prezentare a seriilor statistice:

1. *Serii simple (fără frecvenţă)*:
   a) *cu un număr impar de termeni :* $n = 2p+1$
   b) *cu un număr par de termeni :* $n = 2p$
2. *Serii cu frecvenţă*:
   a) *pe variante ale caracteristicii;*
   b) *pe intervale de variaţie a caracteristicii.*

1a. Să considerăm o serie simplă formată dintr-un număr impar de termeni, ordonaţi în ordine crescătoare sau descrescătoare a caracteristicii:

$$x_1 \le x_2 \le \cdots \le x_p \le x_{p+1} \le x_{p+2} \le \cdots \le x_n$$

sau

$$x_1 \ge x_2 \ge \cdots \ge x_p \ge x_{p+1} \ge x_{p+2} \ge \cdots \ge x_n$$

În general, *locul medianei*, sau numărul de ordine al nivelului caracteristicii egal cu mediana, se determină pe baza formulei:

$$U^{me} = \frac{n+1}{2}$$

pentru $n = 2p+1$ rezultă $U^{me} = p+1$ iar $Me = x_{p+1}$

Se observă că numărul termenilor mai mari decât mediana este egal cu numărul termenilor mai mici decât ea şi egal cu p. Primii p termeni $x_1, x_2, \ldots, x_p$ sunt mai mici decât mediana, iar ultimii p termeni $x_{p+2}, x_{p+3}, \ldots, x_n$ sunt mai mari decât mediana.

1b. Dacă numărul termenilor seriei simple este par, atunci orice nivel al caracteristicii cuprins între x_p şi x_{p+1} poate fi considerat ca o mediană. Se poate afirma că în acest caz nu există o valoare mediană, ci o *zonă mediană* sau un *grup median.*

Locul medianei, în acest caz, este egal cu:

$$U^{me} = \frac{n+1}{2} = \frac{2p+1}{2} = p + \frac{1}{2}$$

Aceasta înseamnă că mediana se găseşte la mijlocul intervalului dintre variantele x_p şi x_{p+1}. În practică, în astfel de cazuri, se admite în mod convenţional

Scanned with OKEN Scanner

că mediana este egală cu media aritmetică simplă a celor două valori centrale ale caracteristicii:

$$U^{me} = \frac{x_p + x_{p+1}}{2}$$

Desigur, acest convenţionalism, utilizat în practică nu derivă din definiţia medianei şi aceasta arată caracterul ei de mărime medie aproximativă.

2a. În cazul distribuţiilor de frecvenţă pe niveluri ale caracteristicii (cazul caracteristicilor discrete), mediana se calculează în următoarele etape:

• Se determină locul medianei:

$$U^{me} = \frac{n+1}{2} = \frac{\sum_{i=1}^{k} n_i + 1}{2} \quad \text{sau} \quad U^{me} = \frac{\sum_{i=1}^{k} f_i + 1}{2}$$

Deoarece, atunci când n este suficient de mare adăugarea unei unităţi la numărătorul unităţii mediane nu influenţează în mod substanţial valoarea U^{me}, în practică, se poate renunţa la adăugarea acesteia, iar unitatea mediană va deveni:

$$U^{me} = \frac{\sum_{i=1}^{k} n_i}{2} \quad \text{sau} \quad U^{me} = \frac{\sum_{i=1}^{k} f_i}{2}$$

• Se determină frecvenţele cumulate:

$$N_i \text{ sau } N_i' \quad \text{sau} \quad F_i \text{ sau } F_i'$$

• Se determină prin comparare:

$$N_i \geq U^{me} \quad \text{sau} \quad N_i' \geq U^{me}$$

sau

$$F_i \geq U^{me} \quad \text{sau} \quad F_i' \geq U^{me}$$

• Mediana este egală cu nivelul caracteristicii x_i care corespunde frecvenţelor cumulate N_i, N_i' sau F_i, F_i' :

$$Me = x_i$$

Scanned with OKEN Scanner

Toate aceste etape de calcul se pot sistematiza cu ajutorul tabelului 6.16

Tabelul 6.16. Determinarea medianei în cazul unei serii cu frecvenţă

X	n_i	N_i
x_1	n_1	N_1
x_2	n_2	N_2
⋮	⋮	⋮
x_i	n_i	N_i
⋮	⋮	⋮
x_k	n_k	N_k
Total	$\sum_1^k n_i$	–

$N_i \geq U^{me}$

$Me = x_i$

Ca şi în cazul precedent, mediana nu este strict determinată, ci ea are aceeaşi valoare ca şi cele n_i variante ale caracteristicii, dar se deosebeşte de cazul precedent prin faptul că acum ea este egală cu o valoare concretă, reală a caracteristicii.

2b. În cazul seriilor pe intervale de variaţie a caracteristicii se pierde individualitatea variantelor variabilei şi de aceea mediana nu se mai poate determina numai pe baza definiţiei, ci este necesară admiterea unei ipoteze complementare. Ipoteza cea mai des considerată în practica calculului medianei este aceea potrivit căreia, în cadrul fiecărui interval, nivelurile individuale ale caracteristicii se repartizează în mod uniform, liniar. Această ipoteză, deşi contravine într-o oarecare măsură definiţiei generale a medianei, este utilizată în practică datorită simplităţii ei, precum şi datorită faptului că erorile ce se comit nu sunt importante din punct de vedere practic.

Tabelul 6.17. Determinarea intervalului median

Caracteristica x_i	Frecvenţa n_i	N_i	N_i'
$x_0 - x_1$	n_1	N_1	N_1'
$x_1 - x_2$	n_2	N_2	N_2'
⋮	⋮	⋮	⋮
$x_{i-1} - x_i$	n_i	N_i	N_i'
⋮	⋮	⋮	⋮
$x_{k-1} - x_k$	n_k	N_k	N_k'
Total	$\sum_1^k n_i$	–	–

Pentru stabilirea formulei de calcul al medianei să considerăm un sistem de axe rectangulare şi să reprezentăm variaţia caracteristicii în cadrul intervalului median printr-un segment de dreaptă AB (fig.6.9).

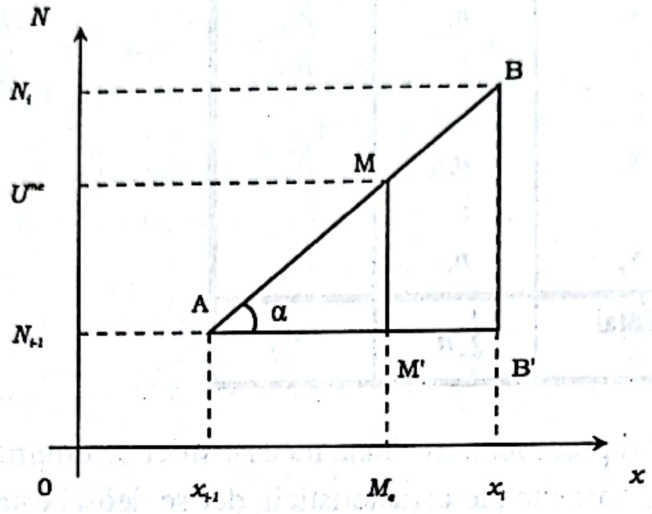


Figura 6.9. Intervalul median, locul şi valoarea medianei

În figura 6.9 s-au format două triunghiuri asemenea $\Delta AMM'$ şi $\Delta ABB'$. Din aceste triunghiuri trebuie să estimăm segmentul AM':

$$tg\alpha = \frac{\overline{MM'}}{\overline{AM'}} = \frac{\overline{BB'}}{\overline{AB'}} \quad \text{iar} \quad \overline{AM'} = \frac{\overline{AB'} \cdot \overline{MM'}}{\overline{BB'}} \tag{6.49}$$

dar $\overline{MM'} = U^{me} - N_{i-1}$ $\qquad$ $\overline{BB'} = N_i - N_{i-1} = n_i$

$\overline{AM'} = Me - x_{i-1}$ $\qquad$ $\overline{AB'} = x_i - x_{i-1} = d$

Substituind aceste valori în relaţia (6.49), rezultă:

$$Me = x_{i-1} + d\frac{U^{me} - N_{i-1}}{n_i} \tag{6.50}$$

unde: x_{i-1} = limita inferioară a intervalului median;
d = mărimea intervalului median;

U^{me} = unitatea mediană;

N_{i-1} = frecvenţa cumulată anterioară intervalului median;

n_i = frecvenţa reală a intervalului median.

Calculul medianei după formula (6.50) necesită parcurgerea următoarelor etape:

1. Determinarea unităţii mediane $U^{me} = \frac{\sum_{i=1}^{k} n_i + 1}{2}$

2. Determinarea frecvenţelor cumulate N_i;
3. Stabilirea intervalului median;
4. Aplicarea formulei (6.50).

Considerăm, pentru exemplificare, seria din tabelul 6.18:

$$U^{me} = \frac{\sum_{i=1}^{9} n_i}{2} = 100 \qquad U^{me} < N_i = 125$$

Din tabelul 6.17. se observă că prima frecvenţă cumulată N_i sau N_i' mai mare sau egală cu U^{me} este egală cu 125. Desigur egalitatea $N_i = N_i'$ se verifică numai în cazuri particulare.

Tabelul 6.18. Distribuţia muncitorilor după nivelul productivităţii muncii

Productivitatea muncii (piese pe schimb)	Numărul muncitorilor (n_i)	N_i	N_i'
255 - 265	5	5	200
265 - 275	10	15	195
275 - 285	20	35	185
285 - 295	40	75	165
295 - 305	50	125	125
305 - 315	30	155	75
315 - 325	25	180	45
325 - 335	10	190	20
335 - 345	10	200	10
Total	200	-	-

Date convenţionale

Intervalul de variaţie a caracteristicii care corespunde frecvenţei cumulate $N_i = 125$, adică (295-305) este intervalul median.

Mediana se poate determina cu ajutorul formulei de interpolare (6.50):

$$Me = 295 + 10\frac{100-75}{50} = 300 \ piese/schimb$$

Aceasta înseamnă că, în medie, un muncitor a produs 300 de piese într-un schimb, sau muncitorul care pe şirul ordonat crescător sau descrescător al nivelului productivităţii muncii, ocupă o poziţie centrală a realizat 300 piese într-un schimb.

6.3.2.1. Determinarea grafică a medianei pentru o serie cu intervale

Pentru a determina mediana pe cale grafică, se reprezintă grafic poligonul frecvenţelor cumulate şi apoi din punctul de ordonată egal cu U^{me} se duce o paralelă la axa *ox* până întâlneşte linia poligonală, iar din acest punct se coboară o perpendiculară pe axa *ox*. Abscisa piciorului perpendicularei este egală cu mediana (fig.6.10).

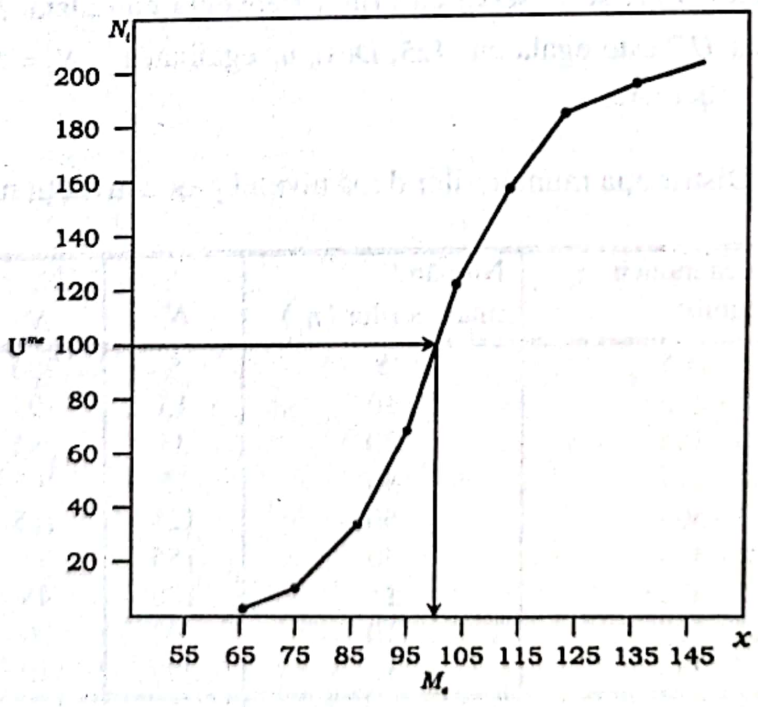


Figura 6.10. Determinarea grafică a medianei

Scanned with OKEN Scanner

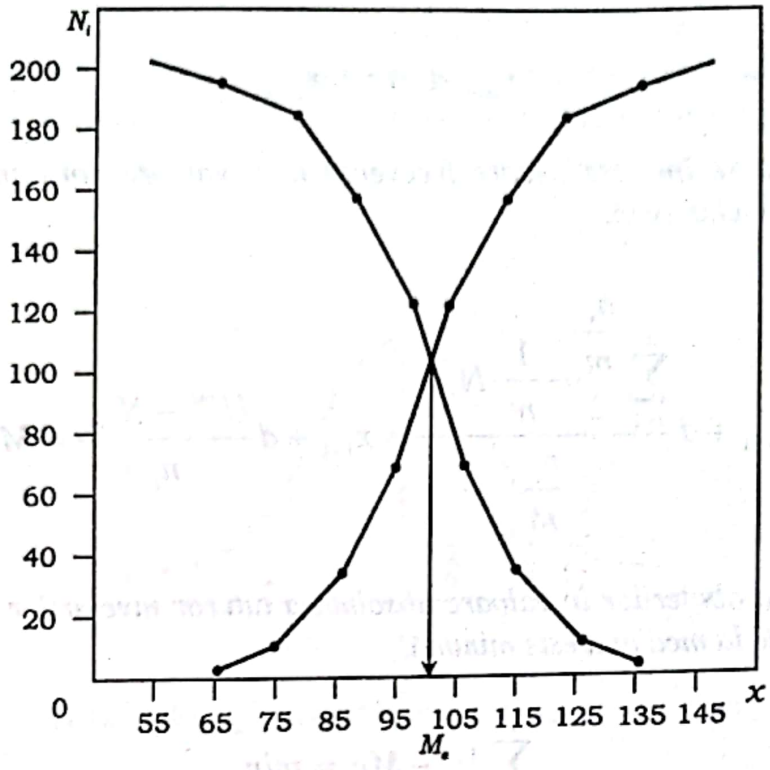


Figura 6.11. Determinarea grafică a medianei pe baza poligoanelor frecvenţelor cumulate

Mediana mai poate fi determinată pe cale grafică, construind pe acelaşi grafic ambele poligoane ale frecvenţelor cumulate N_i şi N_i'. Abscisa punctului de intersecţie al celor două poligoane este egală cu mediana (fig. 6.11).

6.3.2.2. Proprietăţile medianei

Mediana satisface numai parţial *condiţiile Yule-Kendall*, astfel:

a. Mediana ia în considerare toate variantele unei caracteristici, numai sub aspectul rangului şi al numărului lor;

b. Mediana este mai sensibilă decât media aritmetică la fluctuaţiile de sondaj;

c. Mediana nu se pretează la calcule algebrice ulterioare (de exemplu, dacă se iau medianele a două subcolectivităţi pentru o anumită caracteristică, nu se poate determina pe baza acestora mediana întregii colectivităţi);

d. Mediana are o semnificaţie concretă, deoarece ea corespunde în general unei valori observate a caracteristicii.

Proprietăţi:

1. *Mediana este o valoare internă a unei serii statistice:*

Scanned with OKEN Scanner

$$x_{min} < Me < x_{max}$$

2. *Dacă se împarte fiecare frecvenţă la o valoare constantă m mediana seriei rămâne neschimbată.*

$$x_{i-1} + d\frac{\sum_{i=1}^{k}\frac{n_i}{m}}{\frac{n_i}{m}} - \frac{1}{m}\cdot N_{i-1} = x_{i-1} + d\frac{U^{me} - N_{i-1}}{n_i} = Me$$

3. *Suma abaterilor în valoare absolută a tuturor nivelurilor individuale ale caracteristicii de la mediană este minimă:*

$$\sum_{1}^{k} |x_i - Me| = min$$

6.3.2.3. Cuantile

Alături de mediană, există şi alte mărimi de poziţie care împart numărul nivelurilor individuale n ale unei variabile ordonate crescător sau descrescător într-un număr k de părţi egale cu $\frac{n}{k}$. Aceste mărimi sunt cunoscute sub denumirea de *cuantile de ordinul k*. Ele se determină în mod asemănător cu mediana.

Dacă: $n = 2$ cuantila poartă denumirea de *mediană* (Me)
$n = 4$ cuantilele poartă denumirea de *cuartile* (Q_1, Q_2, Q_3)
$n = 10$ cuantilele poartă denumirea de *decile* $(D_1, D_2, \ldots, D_9)$
$n = 100$ cuantilele poartă denumirea de *centile* $(C_1, C_2, \ldots, C_{99})$

Astfel, există trei *cuartile* care împart un şir de valori ale unei variabile în patru părţi, în aşa fel încât fiecare parte să cuprindă acelaşi număr de elemente $\frac{n}{4}$. Prima cuartilă (notată cu Q_1) este acea valoare a unei variabile care împarte o distribuţie în două părţi, astfel încât un sfert din variantele caracteristicii au valori mai mici decât Q_1, iar trei sferturi au valori mai mari decât Q_1. Cuartila a doua (notată cu Q_2) împarte numărul variantelor unei variabile în două părţi egale, astfel încât

Scanned with OKEN Scanner

jumătate dintre variante au valori mai mici decât Q_2 şi jumătate au valori mai mari decât Q_2. Q_2 coincide, deci, cu mediana. Cuartila a treia (notată cu Q_3) împarte numărul termenilor unei serii în două părţi, astfel încât trei sferturi dintre ei au valori mai mici decât Q_3 şi un sfert au valori mai mari decât Q_3. (fig. 6.12)

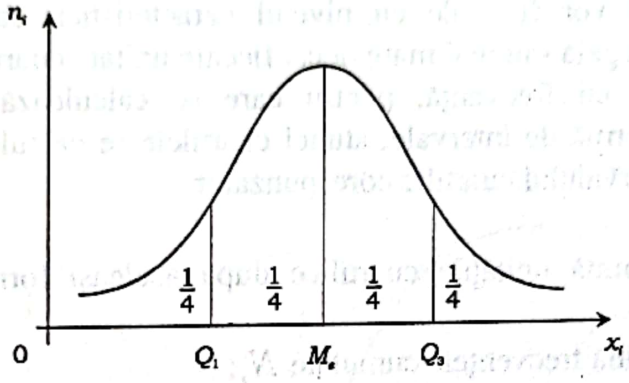


Figura 6.12. Poziţiile cuartilelor într-o distribuţie normală

În calculul cuartilelor se disting două cazuri diferite, după cum seria pentru care se calculează este simplă sau cu frecvenţă.

În cazul unei serii simple , cuartilele se determină astfel:

- se aranjează în ordine crescătoare termenii seriei;
- se determină *locul cuartilei* cu ajutorul *unităţii cuartilice*:

$$U^{Q_1} = \frac{n+1}{4}; \quad U^{Q_2} = \frac{2(n+1)}{4}; \quad U^{Q_3} = \frac{3(n+1)}{4}$$

În cazul unor serii formate dintr-un număr impar de termeni, cuartilele vor fi egale cu variantele caracteristicii care au un număr de ordine egal cu unităţile cuartilice corespunzătoare.

În cazul seriilor formate dintr-un număr par de termeni, unitatea cuartilică va cădea între doi termeni reali ai seriei. În acest caz cuartila corespunzătoare se determină ca o medie aritmetică simplă a celor două niveluri vecine ale caracteristicii.

Dacă cuartilele se calculează pentru o serie cu frecvenţă pe variante ale caracteristicii, atunci se parcurg următoarele etape:

- se determină unităţile cuartilice după formulele:

$$U^{Q_1} = \frac{\sum_{i=1}^{k} n_i + 1}{4}; \quad U^{Q_2} = \frac{2(\sum_{i=1}^{k} n_i + 1)}{4}; \quad U^{Q_3} = \frac{3(\sum_{i=1}^{k} n_i + 1)}{4}$$

- se determină frecvenţele cumulate N_i;
- cuartilele vor fi egale cu nivelul caracteristicii care corespunde unei frecvenţe cumulate egală sau mai mare decât fiecare unitate cuartilică.

Dacă seria cu frecvenţă, pentru care se calculează cuartilele, prezintă caracteristica sub formă de intervale, atunci cuartilele se calculează prin interpolare liniară în cadrul intervalului cuartilic corespunzător.

Astfel:

- se determină unităţile cuartilice după aceleaşi formule ca şi în cazul anterior;
- se determină frecvenţele cumulate N_i;
- se determină intervalele cuartilice, ca fiind egale cu intervalele ce au o frecvenţă cumulată egală sau mai mare decât unitatea cuartilică corespunzătoare;
- se determină prin interpolare fiecare cuartilă utilizând formulele:

$$Q_1 = x_0 + d\frac{U^{Q_1} - N_{ia}}{n_{Q_1}}$$

$$Q_2 = x_0 + d\frac{U^{Q_2} - N_{ia}}{n_{Q_2}} \qquad \text{deci} \quad Q_2 = Me$$

$$Q_3 = x_0 + d\frac{U^{Q_3} - N_{ia}}{n_{Q_3}}$$

In aceste formule:

x_0 = limita inferioară a intervalului cuartilic corespunzător;

d = mărimea intervalului cuartilic corespunzător;

N_{ia} = frecvenţa cumulată anterioară intervalului cuartilic corespunzător;

n_{Q_1} = frecvenţa corespunzătoare intervalului cuartilic al lui Q_1;

n_{Q_2} = frecvenţa corespunzătoare intervalului cuartilic al lui Q_2;

n_{Q_3} = frecvenţa corespunzătoare intervalului cuartilic al lui Q_3;

Scanned with OKEN Scanner

În mod analog se determină decilele şi centilele.

Cele nouă decile împart numărul variantelor unei caracteristici în zece părţi egale cu $\frac{n}{10}$.

Centilele, în număr de 99, împart numărul variantelor unei variabile în 100 părţi egale cu $\frac{n}{100}$.

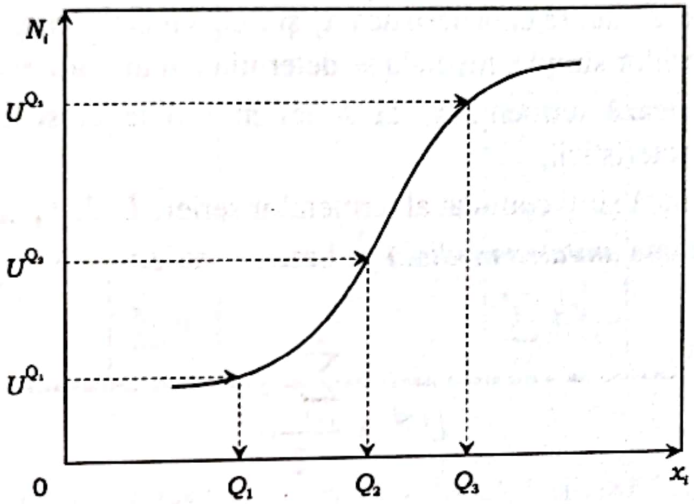


Figura 6.13. Determinarea grafică a cuartilelor

Ca şi mediana cuartilele, decilele şi centilele se pot determina şi pe cale grafică, folosind poligonul frecvenţelor cumulate şi unităţile cuantilice corespunzătoare (fig. 6.13).

6.3.3. Mediala

Mediala, notată cu *Ml* , reprezintă o altă medie de poziţie, asemănătoare prin tehnica de calcul cu mediana.

Nivelul caracteristicii x_i care împarte volumul total al caracteristicii $\sum_{i=1}^{k} x_i n_i$ în două părţi egale se numeşte medială.

Mediala nu se confundă cu mediana, care reprezintă acel nivel al caracteristicii x_i care împarte efectivul total $\sum_{i=1}^{k} n_i$ al unei serii ordonate în două părţi egale.

Din definiţia acestor două mărimi medii de poziţie, rezultă că mediala este totdeauna mai mare decât mediana, calculată pentru aceeaşi distribuţie statistică. Astfel, jumătatea nivelurilor individuale ale caracteristicii inferioare medianei va avea în mod logic un volum total al caracteristicii mai mic decât jumătatea nivelurilor individuale ale caracteristicii care sunt superioare medianei.

Calculul medialei se efectuează în mod similar cu cel al medianei cu deosebirea că, de data aceasta, în locul frecvenţelor cumulate se vor considera produsele cumulate L_i dintre caracteristica x_i şi frecvenţele corespunzătoare n_i.

În cazul seriilor simple, mediala se determină în următoarele etape:

- se ordonează termenii x_i ai seriei în ordine crescătoare a nivelurilor individuale ale caracteristicii;
- se determină şirul cumulat al termenilor seriei L_i de la x_{min} spre x_{max};
- se determină *unitatea medială* pe baza formulei:

$$U^{ml} = \frac{\sum_{i=1}^{k} x_i}{2}$$

- mediala este egală cu acel nivel al caracteristicii x_i care are L_i cumulat egal sau mai mare decât U^{ml}.

În cazul distribuţiilor de freçvenţă pe niveluri ale caracteristicii, mediala se calculează în următoarele etape:

- se determină şirul produselor cumulate $L_i = \sum_{1}^{i} x_i n_i$;
- se determină *unitatea medială* pe baza formulei $U^{ml} = \frac{\sum_{i=1}^{k} x_i n_i}{2}$
- mediala este egală cu nivelul caracteristicii care corespunde produsului cumulat L_i mai mare sau egal cu U^{ml}.

Pentru o serie cu frecvenţă şi cu caracteristica dată sub formă de intervale, mediala se determină prin interpolare liniară în cadrul intervalului medial.

Sistematizarea calculelor preliminare, în vederea stabilirii unei formule de calcul pentru medială, se prezintă în tabelul 6.21. Considerăm că $U^{ml} \leq L_i$. Aceasta înseamnă că mediala va fi cuprinsă în intervalul $x_{i-1} - x_i$ corespunzător produsului

Scanned with OKEN Scanner

cumulat L_i. Presupunem că în cadrul intervalului medial curba produselor cumulate L_i se confundă cu segmentul de dreaptă AB (fig. 6.14).

Tabelul 6.19. Sistematizarea calculelor intermediare pentru determinarea medialei

$x_{i-1} - x_i$	n_i	Mijloacele intervalelor x_i'	$x_i'n_i$	Produse cumulate L_i
$x_0 - x_1$	n_1	x_1'	$x_1'n_1$	L_1
$x_1 - x_2$	n_2	x_2'	$x_2'n_2$	L_2
⋮	⋮	⋮	⋮	⋮
$x_{i-1} - x_i$	n_i	x_i'	$x_i'n_i$	L_i
⋮	⋮	⋮	⋮	⋮
$x_{k-1} - x_k$	n_k	x_k'	$x_k'n_k$	L_k
Total	$\sum_{i=1}^{k} n_i$	–	$\sum_{i=1}^{k} x_i'n_i$	–

S-au format două triunghiuri asemenea $\Delta AMM'$ şi $\Delta ABB'$.

Tangenta unghiului α în cele două triunghiuri va fi:

$$tg\alpha = \frac{\overline{MM'}}{\overline{AM'}} \quad \text{şi} \quad tg\alpha = \frac{\overline{BB'}}{\overline{AB'}} \quad \text{deci} \quad \frac{\overline{MM'}}{\overline{AM'}} = \frac{\overline{BB'}}{\overline{AB'}}$$

ţinând seama de notaţiile din figura 6.14 rezultă:

$$\overline{MM'} = U^{ml} - L_{i-1}$$
$$\overline{AM'} = Ml - x_{i-1}$$
$$\overline{BB'} = L_i - L_{i-1} = x_i'n_i$$
$$\overline{AB'} = x_i - x_{i-1} = d$$

rezultă $$\frac{U^{ml} - L_{i-1}}{Ml - x_{i-1}} = \frac{L_i - L_{i-1}}{x_i - x_{i-1}} = \frac{x_i'n_i}{d}$$

Scanned with OKEN Scanner

de unde $$Ml = x_{i-1} + d\frac{U^{ml} - L_{i-1}}{x_i' n_i} \qquad (6.51)$$

în care: x_{i-1} = limita inferioară a intervalului medial;
d = mărimea intervalului medial;
L_{i-1} = produsul cumulat anterior intervalului medial;
n_i = frecvenţa intervalului medial.

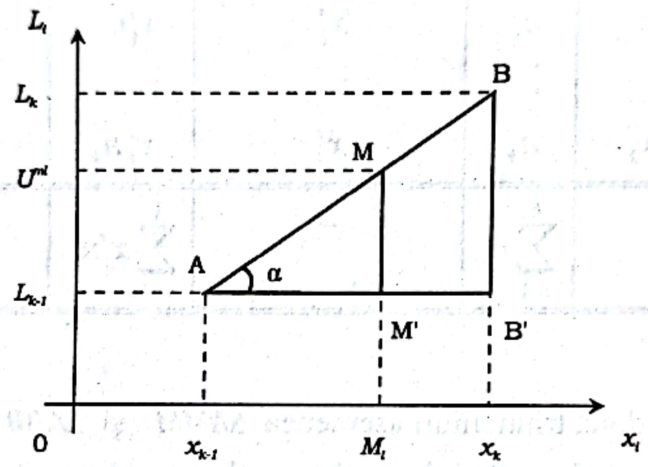


Figura 6.14. Intervalul medial, locul şi valoarea medialei

Pentru exemplificarea calculului medialei, considerăm din nou seria din tabelul 6.18:

$$U^{ml} = \frac{\sum_{i=1}^{k} x_i' n_i}{2} = \frac{60200}{2} = 30100$$

$$U^{ml} = 30100 < 36200$$

deci *intervalul medial* va fi (295-305).

Aplicând formula (6.51) pe baza datelor din tabelul 6.20 rezultă:

Scanned with OKEN Scanner

$$Ml = x_{i-1} + d\frac{\frac{\sum_{i=1}^{k} x_i' n_i}{2} - L_{i-1}}{x_i' n_i} = 295 + 10\frac{30100 - 21200}{300 \cdot 50} = 301 \ piese/zi$$

Aceasta înseamnă că salariatul, care pe şirul ordonat al acestora împarte volumul total al producţiei în două părţi egale, are o productivitate zilnică egală cu 301 piese/zi.

Tabelul 6.20. Calculul medialei

Productivitatea muncii (piese / zi)	Numărul muncitorilor n_i	Mijloacele intervalelor x_i'	$x_i' n_i$	L_i	N_i
255-265	5	260	1300	1300	5
265-275	10	270	2700	4000	15
275-285	20	280	5600	9600	35
285-295	40	290	11600	21200	75
295-305	50	300	15000	36200	125
305-315	30	310	9300	45500	155
315-325	25	320	8000	53500	180
325-335	10	330	3300	56800	190
335-345	10	340	3400	60200	200
Total	200	-	60200	-	-

Ştiind că $Me = 300$ piese/zi, rezultă că se verifică şi relaţia dintre mediană şi medială $Ml > Me$.

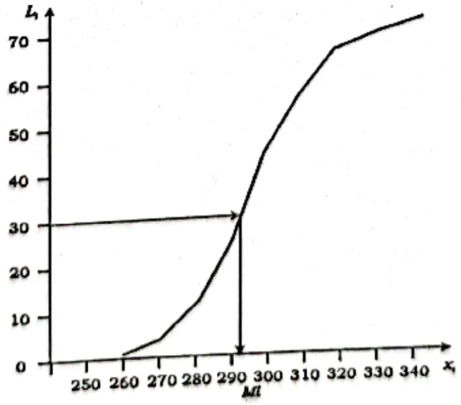


Figura 6.15. Determinarea grafică a medialei

Scanned with OKEN Scanner

Mediala se poate determina şi pe cale grafică. În acest scop considerăm un sistem de axe rectangulare, unde pe abscisă se reprezintă intervalele de variaţie ale caracteristicii, iar pe ordonată produsele cumulate L_i. Se construieşte apoi poligonul produselor cumulate iar din punctul de pe ordonată egal cu U^{mi} se duce o paralelă la axa absciselor până întâlneşte linia poligonală şi din acest punct se coboară o perpendiculară pe axa x. Abscisa piciorului perpendicularei este egală cu mediala (fig 6.15).

Scanned with OKEN Scanner

7. INDICATORII VARIAŢIEI ŞI CONCENTRĂRII

În general, fenomenele şi procesele care constituie obiectul de cercetare al statisticii se prezintă sub formă de colectivităţi sau de fenomene de masă. Fiecare colectivitate se compune dintr-un mare număr de unităţi individuale, care prezintă un anumit nivel al caracteristicii (sau caracteristicilor) de bază supusă cercetării şi care diferă, în general, de la o unitate statistică la alta. Această diversitate este determinată de influenţa unor cauze multiple legate de naşterea şi dezvoltarea fiecărui element component din cadrul colectivităţii.

Pe baza cunoaşterii manifestărilor individuale ale unităţilor statistice, se urmăreşte determinarea unor indicatori sintetici, generalizatori, cu ajutorul cărora să se caracterizeze sub multiple aspecte colectivităţile care fac obiectul ei de cercetare. Aşa după cum am văzut, în capitolul precedent, locul principal în cadrul acestora, este ocupat de indicatorii tendinţei centrale, alături de care se situează *indicatorii împrăştierii* sau ai *variaţiei* nivelurilor individuale ale unei anumite variabile. Indicatorii variaţiei se utilizează pentru caracterizarea colectivităţilor sub aspectul *omogenităţii* lor. Astfel, dintre două colectivităţi de acelaşi fel, formate din acelaşi număr de unităţi, care au acelaşi nivel mediu al unei caracteristici, colectivitatea care are o împrăştiere mai mică are un grad de omogenitate mai ridicat şi invers. Acest fapt se poate ilustra pe cale grafică considerând două distribuţii unimodale simetrice, dar cu boltiri diferite. Curba *a* reprezintă o distribuţie normală cu un grad de concentrare mai mare decât curba *b*. (fig. 7.1)

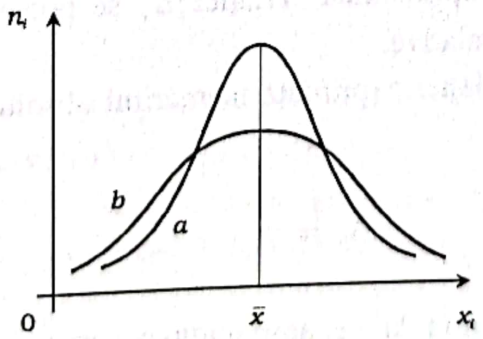


Figura 7.1. Distribuţii unimodale simetrice cu grade diferite de omogenitate

Scanned with OKEN Scanner

Gradul de împrăştiere exprimă şi măsura în care o *medie este reprezentativă* pentru o anumită distribuţie. O valoare mică a indicatorilor de împrăştiere înseamnă o concentrare ridicată a valorilor unei variabile în jurul tendinţei ei centrale, ceea ce dă mediei un caracter reprezentativ şi invers.

Din punct de vedere teoretic, variaţia nivelurilor individuale ale unei caracteristici se poate măsura prin abaterea acestora faţă de o mărime medie oarecare. În practică, însă, indicatorii variaţiei care se bazează pe media aritmetică sunt cel mai des utilizaţi datorită faptului că sunt uşor de calculat şi de interpretat şi se pretează la calcule algebrice ulterioare.

Indicatorii variaţiei se pot calcula şi pe baza medianei, medialei sau a altor mărimi medii.

Indicatorii variaţiei exprimă fie distanţa dintre două niveluri oarecare ale unei variabilei, fie distanţa medie dintre toate nivelurile înregistrate, considerate două câte două, fie distanţa medie dintre nivelurile caracteristicii tuturor unităţilor şi mărimea lor medie.

Indiferent de media care se ia în consideraţie, indicatorii variaţiei sunt de două feluri: *simpli* şi *sintetici*.

Indicatorii simpli ai variaţiei măsoară abaterea fiecărei variante a unei variabile faţă de nivelul ei mediu sau faţă de un alt nivel care are o anumită semnificaţie.

Indicatorii sintetici ai variaţiei măsoară variaţia (împrăştierea) medie a tuturor variantelor unei variabile în jurul nivelului ei mediu.

7.1. Indicatorii simpli ai variaţiei

Un indicator simplu al variaţiei, care oferă o imagine generală asupra dispersiei, este distanţa dintre nivelul maxim şi nivelul minim al variabilei sau *amplitudinea variaţiei*. Amplitudinea variaţiei Δ_x se poate exprima atât în mărimi absolute, cât şi în mărimi relative.

Amplitudinea variaţiei exprimată în mărimi absolute se determină pe baza formulei:

$$\Delta_x = x_{max} - x_{min}$$

Exprimată în mărimi relative, amplitudinea variaţiei se poate calcula:

Scanned with OKEN Scanner

- fie ca raport între x_{max} şi x_{min} : $\Delta'_x = \frac{x_{max}}{x_{min}}$

- fie ca raport între amplitudinea absolută şi nivelul mediu al caracteristicii:

$$\Delta''_x = \frac{x_{max} - x_{min}}{\bar{x}}$$

Amplitudinea variaţiei depinde numai de valorile extreme ale unei variabile şi poate să dea o imagine eronată a intervalului de variaţie, atunci când cel puţin una dintre valorile extreme se abate în mare măsură de la celelalte niveluri individuale ale caracteristicii. Amplitudinea variaţiei nu poate caracteriza în mod corespunzător împrăştierea unor variabile obţinute prin sondaj, în cazul seriilor formate dintr-un număr redus de variante.

Acest inconvenient este înlăturat, în parte, dacă în locul valorilor extreme se vor lua în considerare cuartilele Q_1 şi Q_3, între care se situează 50% dintre valorile individuale ale unei variabile. Indicatorul astfel obţinut poartă denumirea de *interval intercuartilic* şi se determină după formula:

$$d_Q = Q_3 - Q_1$$

Intervalul intercuartilic este împărţit în două părţi egale de către mediană, iar prin raportarea acestuia la 2 se obţine *abaterea intercuartilică*:

$$\Delta_Q = \frac{d_Q}{2} = \frac{Q_3 - Q_1}{2} \qquad (7.1)$$

Abaterea intercuartilică Δ_Q se recomandă a se utiliza în practică, mai ales în cadrul distribuţiilor cu intervale marginale deschise.

Teoretic, abaterea intercuartilică măsoară gradul de împrăştiere a 50% dintre nivelurile individuale ale caracteristicii faţă de mediană. În cazul unor distribuţii simetrice se verifică relaţiile:

$$Me - \Delta_Q = Q_1 \qquad \text{şi} \qquad Me + \Delta_Q = Q_3$$

Scanned with OKEN Scanner

iar în cazul unor distribuţii asimetrice aceste relaţii nu se verifică, adică:

$$Me - \Delta_Q \neq Q_1 \qquad \text{şi} \qquad Me + \Delta_Q \neq Q_3$$

ceea ce înseamnă că intervalul $[Me \pm \Delta_Q]$ nu va include 50% dintre nivelurile individuale ale variabilei. Calculându-se pe baza cuartilelor, *abaterea intercuartilică are o mai mare stabilitate decât amplitudinea variaţiei.*

Abaterea intercuartilică este uşor de calculat, dar prezintă deficienţa că nu ţine seama de valorile variabilei mai mici decât Q_1 şi de cele mai mari decât Q_3. Acest neajuns se înlătură parţial, dacă se va calcula abaterea între decilele 1 şi 9 sau între centilele 10 şi 90, între care se situează 80% dintre variantele unei variabile normal distribuite.

Un alt indicator simplu al variaţiei îl constituie *abaterile nivelurilor individuale* ale unei variabile faţă de nivelul lor mediu, notate cu d_i, şi care se calculează după formula:

$$d_i = x_i - \bar{x}$$

Abaterile d_i se pot exprima şi în mărimi relative prin raportarea lor la nivelul mediu al variabilei, astfel:

$$d_i\% = \frac{d_i}{\bar{x}} \cdot 100 = \frac{x_i - \bar{x}}{\bar{x}} \cdot 100$$

În practică, aceste abateri se utilizează pentru calculul ulterior al indicatorilor sintetici ai variaţiei.

7.2. Indicatorii sintetici ai variaţiei

Indicatorii sintetici ai variaţiei se caracterizează prin faptul că, în calculul lor, se includ abaterile tuturor nivelurilor individuale ale unei variabile faţă de nivelul lor mediu. De aceea, ei caracterizează complet gradul de împrăştiere a variantelor unei variabile în cadrul distribuţiilor.

Din categoria indicatorilor sintetici ai variaţiei fac parte: abaterea medie (zisă şi liniară), dispersia, abaterea medie pătratică şi coeficientul de variaţie.

Scanned with OKEN Scanner

1. *Abaterea medie liniară*, notată cu $\overline{d}$, se determină ca o medie aritmetică (simplă sau ponderată) a abaterilor nivelurilor individuale ale unei variabile faţă de media lor aritmetică, considerate în valoare absolută. Pe baza acestei definiţii abaterea medie liniară se poate calcula cu ajutorul următoarelor formule:

- pentru o serie simplă: $$\overline{d} = \frac{\sum_{i=1}^{n} |x_i - \overline{x}|}{n} \qquad (7.2)$$

- pentru o serie cu frecvenţă:

$$\overline{d} = \frac{\sum_{i=1}^{k} |x_i - \overline{x}| n_i}{\sum_{i=1}^{k} n_i} \quad \text{sau} \quad \overline{d} = \frac{\sum_{i=1}^{k} |x_i - \overline{x}| f_i}{\sum_{i=1}^{k} f_i} \qquad (7.3)$$

Abaterile $|x_i - \overline{x}|$ trebuie considerate în valoare absolută, deoarece dacă s-ar ţine seama de semne, suma lor ar fi egală cu zero, conform proprietăţilor mediei aritmetice.

Abaterea medie liniară se poate calcula şi faţă de mediană.

Formulele de calcul ale acesteia faţă de mediană d' se vor obţine substituind pe $\overline{x}$ cu Me în formulele (7.2) şi (7.3), astfel:

- pentru o serie simplă: $$\overline{d'} = \frac{\sum_{i=1}^{n} |x_i - Me|}{n} \qquad (7.4)$$

- pentru o serie cu frecvenţă:

$$\overline{d'} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{\sum_{i=1}^{k} n_i} \quad \text{sau} \quad \overline{d'} = \frac{\sum_{i=1}^{k} |x_i - Me| f_i}{\sum_{i=1}^{k} f_i} \qquad (7.5)$$

Pe baza proprietăţilor medianei, reţinem faptul că suma abaterilor nivelurilor individuale ale caracteristicii faţă de mediană este mai mică decât în cazul oricărei alte mărimi medii.

Scanned with OKEN Scanner

Aceasta înseamnă că între abaterea medie calculată pe baza mediei aritmetice şi cea calculată pe baza medianei există relaţia: $\overline{d} > \overline{d'}$

Oricare ar fi forma mediei, luate în calcul, abaterea medie liniară prezintă dezavantajul că nu face nici o distincţie între abaterile pozitive şi cele negative.

Să considerăm pentru exemplificare, seria din tabelul 7.1.

Tabelul 7.1. Distribuţia muncitorilor după productivitatea zilnică

Numărul de piese pe zi	Numărul muncitorilor	x_i	$x_i - \overline{x}$	$\lvert x_i - \overline{x}\rvert n_i$	$x_i - Me$	$\lvert x_i - Me\rvert n_i$
255-265	5	260	-41	205	-40	200
265-275	10	270	-31	310	-30	300
275-285	20	280	-21	420	-20	400
285-295	40	290	-11	440	-10	400
295-305	50	300	-1	50	0	0
305-315	30	310	-9	270	10	300
315-325	25	320	-19	475	20	500
325-335	10	330	-29	290	30	300
335-345	10	340	-39	390	40	400
Total	200	-	-	2850	-	2800

Se ştie că: $\overline{x} = 301$ piese/zi iar $Me = 300$ piese/zi.

$$\overline{d} = \frac{\sum_{i=1}^{k} |x_i - \overline{x}| n_i}{\sum_{i=1}^{k} n_i} = \frac{2850}{200} = 14{,}25 \ piese/zi$$

$\overline{d} = 14{,}25$ înseamnă că, în medie, un muncitor din grupul considerat se abate de la productivitatea medie cu $\pm 14{,}25$ piese.

$$\overline{d'} = \frac{\sum_{i=1}^{k} |x_i - Me| n_i}{\sum_{i=1}^{k} n_i} = \frac{2800}{200} = 14 \ piese/zi$$

Scanned with OKEN Scanner

Se observă că $\overline{d'} < \overline{d}$ dar diferenţa între ele este foarte mică deoarece distribuţia este foarte apropiată de cea normală:

$$\bar{x} - Me = 301 - 300 = 1 \text{ piese/zi.}$$

Oricare ar fi natura mediei luate în calcul, abaterea medie liniară prezintă dezavantajul că nu face nici o distincţie între abaterile pozitive şi cele negative ale variantelor unei caracteristici deşi în practică acestea au semnificaţii cu totul deosebite.

2. *Dispersia*, notată cu σ^2, este un alt indicator sintetic al variaţiei mai mult cu o semnificaţie de calcul dar cu o importanţă deosebită în analiza influenţei factorilor care acţionează în cadrul colectivităţilor. Dispersia se calculează ca o medie aritmetică a pătratelor abaterilor nivelurilor individuale ale unei variabile de la nivelul lor mediu (media aritmetică). Deoarece abaterile la puterea întâia au semne diferite şi prin însumare, la nivelul seriei, se compensează reciproc este necesar ca acestea să se ridice la pătrat înainte de însumare. Pe baza acestei definiţii formulele de calcul ale dispersiei sunt:

- pentru o serie simplă:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} \tag{7.6}$$

- pentru o serie cu frecvenţă:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} \quad \text{sau} \quad \sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i} \tag{7.7}$$

3. *Abaterea medie pătratică*, numită şi *abatere standard* sau *abatere tip* se notează cu σ şi se calculează prin extragerea rădăcinii pătrate din dispersie. Abaterea medie pătratică reprezintă media pătratică a abaterilor nivelurilor individuale ale caracteristicii faţă de media lor.

Formulele de calcul ale abaterii standard sunt următoarele:

Scanned with OKEN Scanner

- pentru o serie simplă: $$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \qquad (7.8)$$

- pentru o serie cu frecvenţă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i}} \quad \text{sau} \quad \sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i}} \qquad (7.9)$$

În exemplul din tabelul 7.1, pentru determinarea abaterii medi pătratice calculele se sistematizează cu ajutorul tabelului 7.2.

În cazul utilizării frecvenţelor exprimate în mărimi absolute, abaterea medie pătratică a seriei se calculează pe baza formulei (7.9), astfel:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i}} = \sqrt{\frac{66800}{200}} = \sqrt{334} = 18{,}27 \ \textit{piese/zi}$$

Acelaşi rezultat se obţine dacă în locul frecvenţelor absolute se vor utiliza frecvenţele relative (ultimele două coloane ale tabelului 7.2).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 f_i}{\sum_{i=1}^{k} f_i}} = \frac{33400}{100} = 18{,}27 \ \textit{piese/zi}$$

$\sigma = 18{,}27$ piese/zi înseamnă că, în medie, productivitatea zilnică a muncii unui muncitor se abate de la nivelul mediu al grupului cu $\pm 18{,}27$ piese/zi.

Scanned with OKEN Scanner

Tabelul 7.2. Determinarea abaterii medii pătratice

x_i	n_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$f_i(\%)$	$(x_i - \bar{x})^2 f_i$
260	5	-41	1681	8405	2,5	4202,5
270	10	-31	961	9610	5,0	4805,0
280	20	-21	441	8820	10,0	4410,0
290	40	-11	121	4840	20,0	2420,0
300	50	-1	1	50	25,0	25,0
310	30	9	81	2430	15,0	1215,0
320	25	19	361	9025	12,0	4512,5
330	10	29	841	8410	5,0	4205,0
340	10	39	1521	15210	5,0	7605,0
Total	200			66800	100,0	33400,0

În cazul unei distribuţii normale există anumite relaţii bine definite între abaterea intercuartilică, abaterea medie şi abaterea medie pătratică şi anume:

• *abaterea intercuartilică este egală cu două treimi din abaterea medie pătratică*:

$$\Delta_Q = 0{,}6745\,\sigma \tag{7.10}$$

• *abaterea medie liniară este egală cu patru cincimi din abaterea medie pătratică*:

$$\bar{d} = 0{,}7979\,\sigma \tag{7.11}$$

O trăsătură importantă a abaterii medii pătratice, ca indicator sintetic al împrăştierii, este faptul că, în cazul unei distribuţii normale, o proporţie fixă a variantelor unei variabile se cuprinde în intervalul $[\bar{x} \pm k\sigma]$ unde $k \in \Re$. Astfel, 68,27% dintre variantele unei variabile normal distribuite (fig.7.2) se cuprind în intervalul $[\bar{x} \pm \sigma]$; 95,45% în intervalul $[\bar{x} \pm 2\sigma]$ şi 99,73% în intervalul $[\bar{x} \pm 3\sigma]$.

Din punct de vedere practic, este util de reţinut faptul că 95% din variantele unei variabile se includ în intervalul $[\bar{x} \pm 1{,}96\sigma]$ şi, evident, în afara acestui interval se vor afla 5%, iar în intervalul $[\bar{x} \pm 2{,}58\sigma]$ se cuprind 99% dintre variantele unei variabile şi numai 1% vor rămâne în afara lui. Este de remarcat şi

Scanned with OKEN Scanner

faptul că aceste proporţii se realizează cu aproximaţie şi în cazul distribuţiilor unimodale care nu sunt normale.

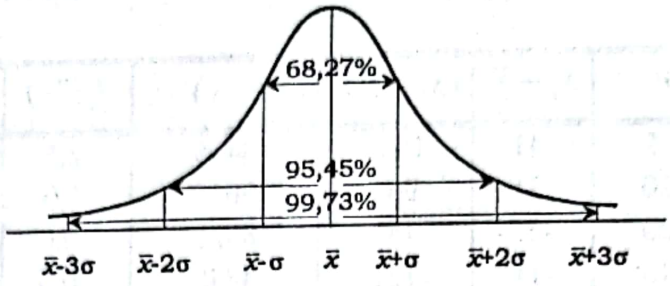


Figura 7.2. Proporţia variantelor unei variabile normal distribuite cuprinse în intervalul $x \pm 3\sigma$

În practică, dacă termenii unei serii simple variază aproximativ în progresie geometrică, atunci în locul dispersiei obişnuite se poate utiliza aşa-numita *dispersie geometrică* σ_g^2 calculată pe baza formulei:

$$\sigma_g^2 = \frac{\sum_{i=1}^{n}(\lg x_i - \lg \bar{x}_g)^2}{n} \quad \text{sau} \quad \sigma_g^2 = \frac{\sum_{i=1}^{k}(\lg x_i - \lg \bar{x}_g)^2 n_i}{\sum_{i=1}^{k} n_i}$$

Extrăgând rădăcina pătrată din dispersia geometrică se obţine *abaterea medie pătratică geometrică* σ_g:

$$\sigma_g = \sqrt{\frac{\sum_{i=1}^{n}(\lg x_i - \lg \bar{x}_g)^2}{n}} \quad \text{sau} \quad \sigma_g = \sqrt{\frac{\sum_{i=1}^{k}(\lg x_i - \lg \bar{x}_g)^2 n_i}{\sum_{i=1}^{k} n_i}}$$

Abaterea medie liniară, dispersia şi abaterea medie pătratică se exprimă în aceleaşi unităţi de măsură ca şi variabila luată în calcul. De aceea, aceşti indicatori se pot utiliza pentru caracterizarea gradului de împrăştiere în cadrul oricărei distribuţii şi pentru compararea împrăştierii unor distribuţii diferite, de aceeaşi natură şi cu caracteristica exprimată în aceleaşi unităţi de măsură. Ei nu pot fi folosiţi pentru

Scanned with OKEN Scanner

caracterizarea împrăştierii unor colectivităţi cu caracteristicile exprimate în unităţi de măsură diferite, ca şi pentru colectivităţile care, deşi au caracteristicile exprimate în aceleaşi unităţi de măsură, diferă prin natura lor.

În aceste situaţii se utilizează expresia relativă (standardizată) a abaterii medii pătratice, numită şi *coeficient de variaţie*.

4. *Coeficientul de variaţie*, notat cu v_σ, se calculează ca raport între abaterea medie pătratică şi nivelul mediu al unei variabile, iar pentru a se interpreta mai uşor se exprimă în procente:

$$v_\sigma = \frac{\sigma}{\bar{x}} \cdot 100 \qquad (7.12)$$

Coeficientul de variaţie arată în ce raport se găseşte împrăştierea unei variabile, măsurată prin abaterea medie pătratică, faţă de nivelul ei mediu. Datorită faptului că acest indicator sintetic al variaţiei se exprimă în procente, făcând deci abstracţie de unităţile de măsură concrete ale caracteristicilor, el poate fi utilizat pentru compararea gradului de împrăştiere al unor distribuţii care au caracteristica exprimată în unităţi de măsură diferite; iar întrucât el reprezintă expresia relativă a abaterii medii pătratice, se poate folosi pentru compararea împrăştierii unor distribuţii care diferă foarte mult prin natura lor (mediile sunt foarte diferite), deşi variabilele lor se exprimă în aceleaşi unităţi de măsură.

Coeficientul de variaţie se poate calcula şi prin raportarea abaterii medii liniare la media aritmetică a distribuţiei, astfel:

$$v_d = \frac{\bar{d}}{\bar{x}} \cdot 100 \qquad (7.13)$$

Caracterizarea variaţiei unei variabile se poate realiza şi cu ajutorul *cuartilelor*.

Intervalul interquartilic d_Q, calculat ca o diferenţă între cuartila a treia şi prima cuartilă: $Q_3 - Q_1$ poate caracteriza gradul de împrăştiere (de dispersie) pentru o anumită variabilă:

$$d_Q = Q_3 - Q_1$$

d_Q se exprimă în aceeaşi unitate de măsură ca şi variabila pentru care se calculează, fapt care-i reduce aria de aplicabilitate; el nu se pretează la compararea

Scanned with OKEN Scanner

unor serii diferite care au parametrii tendinţei centrale diferiţi şi amplitudini diferite ale variaţiei sau caracteristica exprimată în unităţi de măsură diferite.

În acest din urmă caz, se recurge la expresia relativă d'_Q a lui d_Q prin raportarea intervalului interquartilic, exprimat în mărimi absolute, la amplitudinea variaţiei a_x, ori la media aritmetică $\bar{x}$ sau la mediană Me, astfel:

$$d'_Q = \frac{d_Q}{a_x} = \frac{Q_3 - Q_1}{a_x}$$

$$d''_Q = \frac{d_Q}{\bar{x}} = \frac{Q_3 - Q_1}{\bar{x}}$$

$$d'''_Q = \frac{d_Q}{Me} = \frac{Q_3 - Q_1}{Me}$$

Toţi aceşti indicatori relativi ai dispersiei sunt mărimi scalare şi, deci, au o arie largă de aplicabilitate.

d'_Q este un indicator care caracterizează gradul de concentrare a variantelor unei variabile în jumătatea din centrul seriei şi arată proporţia ocupată de d_Q din a_x.

Acest fapt se poate ilustra sugestiv pe cale grafică (fig. 7.3)

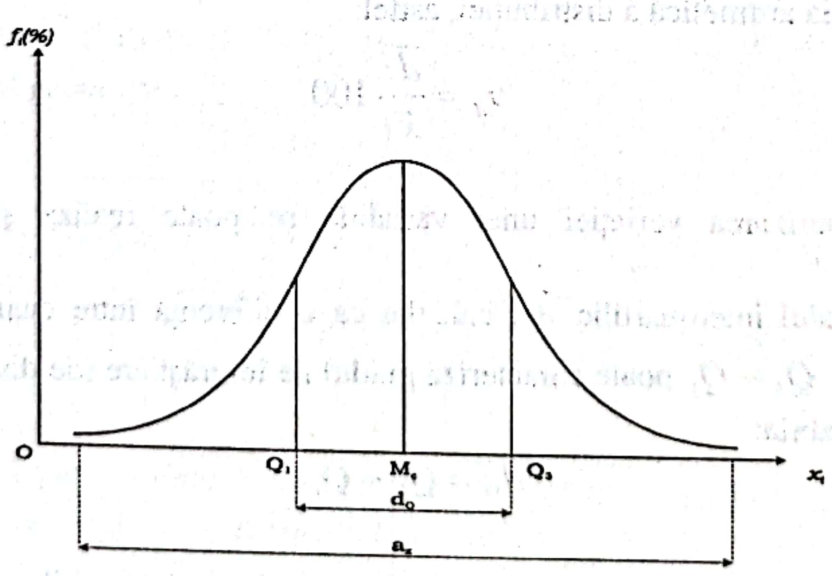


Figura 7. 3. Amplitudinea variaţiei şi abaterea interquartilică

Scanned with OKEN Scanner

Astfel, absenţa concentrării variantelor unei variabile în centrul unei serii va fi dată de $d'_Q = 0{,}5$ ceea ce înseamnă că 50% dintre variantele unei variabile se dispersează în 50% din amplitudinea variaţiei.

$d'_Q > 0{,}5$ va arăta că în cadrul seriei concentrarea este mai slabă.

Cu cât d'_Q este mai mic decât 0,5 cu atât concentrarea variantelor variabilei în centrul seriei va fi mai puternică şi invers.

Un alt indicator relativ al variaţiei este *coeficientul de variaţie intercuartilică*, care se calculează prin raportarea abaterii intercuartilice la valoarea medianei unei distribuţii:

$$v_{\Delta_Q} = \frac{\Delta_Q}{Me} \cdot 100 \qquad (7.14)$$

sau

$$v'_{\Delta_Q} = \frac{\frac{Q_3 - Q_1}{2}}{\frac{Q_3 + Q_1}{2}} = \frac{Q_3 - Q_1}{Q_3 + Q_1} \qquad (7.15)$$

Din (7.15) se observă că v'_{Δ_Q} variază între două limite precise şi anume:

$$-1 \le v_{\Delta_Q} \le 1$$

În exemplul anterior, coeficientul de variaţie ia următoarele valori:

$$v_\sigma = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{18{,}27}{301} \cdot 100 = 6{,}07\%$$

$$v_d = \frac{\bar{d}}{\bar{x}} \cdot 100 = \frac{14{,}25}{301} \cdot 100 = 4{,}73\%$$

$$v_{\Delta_Q} = \frac{\Delta_Q}{Me} \cdot 100 = \frac{12{,}265}{300} \cdot 100 = 4{,}09\%$$

$$v'_{\Delta_Q} = \frac{Q_3 - Q_1}{Q_3 + Q_1} = \frac{313{,}33 - 288{,}75}{313{,}33 + 288{,}75} = \frac{24{,}58}{602{,}08} = 0{,}04$$

Scanned with OKEN Scanner

$\nu_\sigma = 6{,}07\%$ înseamnă că nivelurile individuale ale caracteristicii se abat în medie de la media lor aritmetică cu 6,07% în ambele sensuri.

$\nu_{\Delta_Q} = 4{,}09\%$ arată că variantele caracteristicii se abat în medie de la nivelul median al caracteristicii cu 4,09% în ambele sensuri.

7.2.1. Proprietăţile dispersiei şi calculul ei simplificat

Dispersia prezintă o serie de proprietăţi a căror cunoaştere este utilă pentru o mai bună înţelegere a conţinutului acestui indicator, cât şi pentru efectuarea calculului ei simplificat. Cele mai importante proprietăţi ale dispersiei vor fi tratate în continuare.

1. *Dispersia unui şir de valori constante este egală cu zero*, întrucât toate variantele caracteristicii sunt egale între ele, iar valoarea lor comună este media lor aritmetică:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = 0 \qquad \text{sau} \qquad \sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = 0$$

unde $x_i - \bar{x} = 0$ deoarece $x_i = \bar{x}$.

2. *Dispersia unei distribuţii este egală cu diferenţa dintre media pătratelor tuturor variantelor caracteristicii* $\overline{x^2}$ *şi pătratul mediei* $\bar{x}^2$.

Pentru demonstrarea acestei proprietăţI, se poate porni de la formula generală a dispersiei (7.7) în care se efectuiază calculele cerute de semne:

- în cazul unei serii simple:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\frac{(\sum_{i=1}^{n} x_i)^2}{n^2}}{n} =$$

$$= \frac{\sum_{i=1}^{n} x_i^2}{n} - 2\frac{(\sum_{i=1}^{n} x_i)^2}{n^2} + \frac{(\sum_{i=1}^{n} x_i)^2}{n^2} = \frac{\sum_{i=1}^{n} x_i^2}{n} - \frac{(\sum_{i=1}^{n} x_i)^2}{n} = \overline{x^2} - \bar{x}^2$$

Scanned with OKEN Scanner

• în cazul unei serii cu frecvenţe:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i-\bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}(x_i^2 - 2x_i\bar{x}+\bar{x}^2)n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i^2 n_i - 2\bar{x}\sum_{i=1}^{k} x_i n_i + \bar{x}^2 \sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i} =$$

$$= \frac{\sum_{i=1}^{k} x_i^2 n_i - 2\frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i}\sum_{i=1}^{k} x_i n_i + \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{\left(\sum_{i=1}^{k} n_i\right)^2}\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i} =$$

$$\frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} - 2\frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{\left(\sum_{i=1}^{k} n_i\right)^2} + \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{\left(\sum_{i=1}^{k} n_i\right)^2} = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} - \frac{\left(\sum_{i=1}^{k} x_i n_i\right)^2}{\left(\sum_{i=1}^{k} n_i\right)^2} = \overline{x^2} - \bar{x}^2$$

deci

$$\sigma^2 = \overline{x^2} - \bar{x}^2 \qquad (7.16)$$

Relaţia (7.16) exprimă faptul că *dispersia este egală cu momentul obişnuit de ordinul doi minus momentul obişnuit de ordinul întâi la pătrat.*

Atât pentru distribuţiile simple, cât şi pentru cele cu frecvenţă calculul dispersiei se efectuează pe baza aceleiaşi formule (7.16), ţinând seama de faptul că în primul caz mediile respective vor fi medii aritmetice simple, iar în cel de-al doilea, medii ponderate.

În mod corespunzător, abaterea medie pătratică se va calcula după formula:

Scanned with OKEN Scanner

$$\sigma = \sqrt{\sigma^2} = \sqrt{\overline{x^2} - \overline{x}^2} \qquad (7.17)$$

Aplicarea în practică a formulelor (7.16) și (7.17) duce la o simplificare a calculelor, în sensul că nu mai necesită determinarea prealabilă a abaterilor de la medie $d_i = x_i - \overline{x}$.

Calculul simplificat al abaterii medii pătratice pe baza formulei (7.17), în exemplul anterior, se sistematizează cu ajutorul tabelului 7.3.

$$\overline{x^2} = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{18187000}{200} = 90935$$

Tabelul 7.3. Calculul simplificat al dispersiei

x_i	n_i	$x_i n_i$	x_i^2	$x_i^2 n_i$
260	5	1300	61600	308000
270	10	2700	72900	729000
280	20	5600	78400	1568000
290	40	11600	84100	3364000
300	50	1500	90000	4500000
310	30	9300	96100	2883000
320	25	8000	102400	2560000
330	10	3300	108900	1089000
340	10	3400	115600	1156000
Total	200	60200	-	18187000

$$\overline{x} = \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} = \frac{60200}{200} = 301 \ piese/zi$$

Scanned with OKEN Scanner

şi $$\sigma^2 = \overline{x^2} - \bar{x}^2 = 90935 - 90601 = 334$$

iar $$\sigma = \sqrt{334} = 18{,}27 \ piese/zi$$

Se observă că s-a obţinut acelaşi rezultat al abaterii medii pătratice ca şi în cazul utilizării formulei de bază.

3. *Dacă se adună sau se scade o valoare constantă c din fiecare variantă a unei caracteristici atunci dispersia rămâne neschimbată.*

• În cazul unei serii simple:

$$\frac{\sum_{i=1}^{n}[(x_i \pm c)-(\bar{x} \pm c)]^2}{n} = \frac{\sum_{i=1}^{n}(x_i \pm c)^2}{n} - 2(\bar{x} \pm c)\frac{\sum_{i=1}^{n}(x_i \pm c)}{n} + \frac{n(\bar{x} \pm c)^2}{n} =$$

$$\frac{\sum_{i=1}^{n}x_i^2}{n} \pm 2c\frac{\sum_{i=1}^{n}x_i}{n} + \frac{nc^2}{n} - 2(\bar{x} \pm c)(\bar{x} \pm c) + (\bar{x} \pm c)^2 =$$

$$= \overline{x^2} \pm 2c\bar{x} + c^2 - \bar{x}^2 \mp 2c\bar{x} - c^2 = \overline{x^2} - \bar{x}^2 = \sigma^2$$

deci $$\sigma^2 = \frac{\sum_{i=1}^{n}[(x_i \pm c)-(\bar{x} \pm c)]^2}{n} \qquad (7.18)$$

• În cazul unei serii cu frecvenţă:

$$\sigma^2 = \frac{\sum_{i=1}^{k}[(x_i \pm c)-(\bar{x} \pm c)]^2 n_i}{\sum_{i=1}^{k} n_i} \qquad (7.19)$$

4. *Dacă se înmulţesc toate variantele unei caracteristici cu o constantă m atunci dispersia se măreşte de m^2 ori.*

Scanned with OKEN Scanner

• În cazul unei serii simple:

$$\frac{\sum_{i=1}^{n}(x_i m - m\bar{x})^2}{n} = \frac{m^2\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = m^2\sigma$$

de unde $$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i m - m\bar{x})^2}{nm^2} \qquad (7.20)$$

• În cazul unei serii cu frecvenţă:

$$\frac{\sum_{i=1}^{k}(mx_i - m\bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{m^2\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = m^2\sigma^2$$

de unde $$\sigma^2 = \frac{\sum_{i=1}^{k}(mx_i - m\bar{x})^2 n_i}{m^2\sum_{i=1}^{k} n_i} \qquad (7.21)$$

5. *Dispersia calculată din abaterile variantelor caracteristicii faţă de o constantă a , aleasă arbitrar, este mai mare decât dispersia calculată faţă de media lor aritmetică cu pătratul diferenţei dintre media aritmetică şi constanta a.*

• În cazul unei serii simple:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(x_i - a)^2}{n} - (\bar{x} - a)^2 \qquad a \neq \bar{x} \qquad (7.22)$$

• În cazul unei serii cu frecvenţă:

Scanned with OKEN Scanner

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - a)^2 n_i}{\sum_{i=1}^{k} n_i} - (\bar{x} - a)^2 \qquad a \neq \bar{x} \tag{7.23}$$

Astfel, dacă se efectuează calculele în relaţia (7.23), rezultă:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - a)^2 n_i}{\sum_{i=1}^{k} n_i} - (\bar{x} - a)^2 = \frac{\sum_{i=1}^{k} x_i^2 n_i - 2a\sum_{i=1}^{k} x_i n_i + a^2 \sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i} -$$

$$-(\bar{x}^2 - 2a\bar{x} + a^2) = \frac{\sum_{i=1}^{k} x_i^2 n_i}{\sum_{i=1}^{k} n_i} - 2a\bar{x} + a^2 - \bar{x}^2 + 2a\bar{x} - a^2 = \overline{x^2} - \bar{x}^2$$

Pe baza acestei proprietăţi a dispersiei, abaterea medie pătratică se poate calcula cu ajutorul formulelor:

- pentru o serie simplă: $\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - a)^2}{n} - (\bar{x} - a)^2}$

- pentru o serie cu frecvenţă: $\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - a)^2 n_i}{\sum_{i=1}^{k} n_i} - (\bar{x} - a)^2}$

Reluând exemplul anterior, calculul simplificat al dispersiei se poate sistematiza cu ajutorul tabelului 7.4.

Scanned with OKEN Scanner

Tabelul 7.4. Calculul simplificat al dispersiei

x_i	n_i	$x_i - a$	$(x_i - a)^2$	$(x_i - a)^2 n_i$
260	5	-40	1600	8000
270	10	-30	900	9000
280	20	-20	400	8000
290	40	-10	100	4000
300	50	0	0	0
310	30	10	100	3000
320	25	20	400	10000
330	10	30	900	9000
340	10	40	1600	16000
Total	200	-	-	67000

Considerăm $a = 300$, adică o valoare a caracteristicii din centrul distribuţiei.

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - a)^2 n_i}{\sum_{i=1}^{k} n_i} - (\bar{x} - a)^2 = \frac{67000}{200} - (301 - 300)^2 = 334$$

6. *Dacă fiecare abatere d_i a unei variabile se micşorează de b ori, atunci dispersia se micşorează de b^2 ori.*

• Pentru o serie simplă:

$$\frac{\sum_{i=1}^{n}\left(\frac{x_i - \bar{x}}{b}\right)^2}{n} = \frac{\frac{1}{b^2}\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sigma^2}{b^2} \qquad (7.24)$$

• Pentru o serie cu frecvenţă:

Scanned with OKEN Scanner

$$\frac{\sum_{i=1}^{k}\left(\frac{x_i-\bar{x}}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{\frac{1}{b^2}\sum_{i=1}^{k}(x_i-\bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{\sigma^2}{b^2} \qquad (7.25)$$

Din relaţiile (7.24) şi (7.25) rezultă următoarele formule de calcul simplificat al abaterii medii pătratice:

- pentru o serie simplă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{b}\right)^2}{n}\cdot b^2} \qquad (7.26)$$

- pentru o serie cu frecvenţă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{x_i-\bar{x}}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i}\cdot b^2} \qquad (7.27)$$

Calculele necesare pentru aplicarea în practică a relaţiei (7.27) se sistematizează cu ajutorul tabelului 7.5.

În formula (7.27) avem:

$$\frac{x_i-\bar{x}}{b} = \frac{x_i}{b} - \frac{\bar{x}}{b} = \frac{x_i}{10} - \frac{301}{10}$$

întrucât am considerat $b = 10$

Scanned with OKEN Scanner

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{x_i - \bar{x}}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i} \cdot b^2} = \sqrt{\frac{668}{200} \cdot 100} = \sqrt{334} = 18{,}27 \quad piese/zi$$

Tabelul 7.5. Calculul simplificat al abaterii medii pătratice

x_i	n_i	$\frac{x_i}{k}$	$\frac{x_i}{k} - 30{,}1$	$\left(\frac{x_i}{k} - 30{,}1\right)^2$	$\left(\frac{x_i}{k} - 30{,}1\right)^2 n_i$
260	5	26	-4,1	16,81	84,05
270	10	27	-3,1	9,61	96,10
280	20	28	-2,1	4,41	88,20
290	40	29	-1,1	1,21	48,40
300	50	30	-0,1	0,01	0,50
310	30	31	0,9	0,81	24,30
320	25	32	1,9	3,61	90,25
330	10	33	2,9	8,41	84,10
340	10	34	3,9	15,21	152,10
Total	200				668,00

7. Dacă se împarte fiecare frecvenţă a caracteristicii unei distribuţii printr-o constantă c , atunci dispersia rămâne neschimbată.

Astfel:

$$\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \frac{n_i}{c}}{\sum_{i=1}^{k}\frac{n_i}{c}} = \frac{\frac{1}{c}\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\frac{1}{c}\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \sigma^2 \qquad (7.28)$$

Pe baza relaţiei (7.28) formula de calcul simplificat a abaterii medii pătratice va fi:

Scanned with OKEN Scanner

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 \frac{n_i}{c}}{\sum_{i=1}^{k}\frac{n_i}{c}}} \qquad (7.29)$$

O simplificare mai mare a calculelor dispersiei se obţine dacă se utilizează concomitent mai multe proprietăţi ale ei. Astfel, în baza proprietăţilor 5 şi 6, dispersia se va calcula cu ajutorul formulei:

- pentru o serie simplă:

$$\sigma^2 = \frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{b}\right)^2}{n} b^2 - (\bar{x} - a)^2 \qquad (7.30)$$

- pentru o serie cu frecvenţă:

$$\sigma^2 = \frac{\sum_{i=1}^{k}\left(\frac{x_i - a}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i} b^2 - (\bar{x} - a)^2 \qquad (7.31)$$

Iar abaterea medie pătratică va fi:

- pentru o serie simplă:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{x_i - a}{b}\right)^2}{n} b^2 - (\bar{x} - a)^2} \qquad (7.32)$$

- pentru o serie cu frecvenţă:

Scanned with OKEN Scanner

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{x_i - a}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i} b^2 - (\bar{x} - a)^2} \qquad (7.33)$$

Utilizând proprietăţile 5, 6 şi 7 se ajunge la o şi mai mare simplificare a calculelor dispersiei şi abaterii medii pătratice, deşi în aparenţă formulele sunt mai complicate. Formula de calcul a dispersiei va fi:

$$\sigma^2 = \frac{\sum_{i=1}^{k}\left(\frac{x_i - a}{b}\right)^2 \frac{n_i}{c}}{\sum_{i=1}^{k} \frac{n_i}{c}} b^2 - (\bar{x} - a)^2 \qquad (7.34)$$

iar abaterea medie pătratică:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{x_i - a}{b}\right)^2 \frac{n_i}{c}}{\sum_{i=1}^{k} \frac{n_i}{c}} b^2 - (\bar{x} - a)^2} \qquad (7.35)$$

8. *Dispersia unei variabile X înmulţită cu o constantă k este egală cu produsul dintre dispersia variabilei X şi pătratul constantei k* .Astfel:

- în cazul seriilor simple:

$$X: \quad (x_1, \quad x_2, \quad \ldots, \quad x_i, \quad \ldots, \quad x_n)$$

avem $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$ şi $\sigma^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$

$$X' = kX$$

Scanned with OKEN Scanner

$$X': \quad (kx_1, \quad kx_2, \quad \ldots, \quad kx_i, \quad \ldots, \quad kx_n)$$

$$\sigma^2_{x'} = \frac{\sum_{i=1}^{n}(kx_i - \bar{x}')^2}{n} \quad \text{unde} \quad \bar{x}' = k\bar{x} \quad \text{şi} \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\sigma^2_{x'} = \frac{\sum_{i=1}^{n}(kx_i - k\bar{x})^2}{n} = \frac{k^2\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = k^2\sigma^2 \qquad (7.36)$$

• în cazul seriilor cu frecvenţă:

$$X: \quad \begin{pmatrix} x_1, & x_2, & \ldots, & x_i, & \ldots, & x_m \\ n_1, & n_2, & \ldots, & n_i, & \ldots, & n_m \end{pmatrix}$$

avem $\bar{x} = \dfrac{\sum_{i=1}^{m} x_i n_i}{\sum_{i=1}^{m} n_i}$ şi $\sigma^2_x = \dfrac{\sum_{i=1}^{m}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{m} n_i}$

$$X': \quad \begin{pmatrix} kx_1, & kx_2, & \ldots, & kx_i, & \ldots, & kx_m \\ n_1, & n_2, & \ldots, & n_i, & \ldots, & n_m \end{pmatrix}$$

$$\bar{x}' = \frac{\sum_{i=1}^{m} kx_i n_i}{\sum_{i=1}^{m} n_i} = k\bar{x}$$

şi

Scanned with OKEN Scanner

$$\sigma^2_{x'} = \frac{\sum_{i=1}^{m}(kx_i - \bar{x}')^2 n_i}{\sum_{i=1}^{m} n_i} = \frac{\sum_{i=1}^{m}(kx_i - k\bar{x})^2 n_i}{\sum_{i=1}^{m} n_i} = k^2\sigma^2_x$$

9. *Dispersia sumei a două variabile* $Z = X + Y$ *este egală cu suma dispersiilor celor două variabile, la care se adaugă de două ori covarianța lor.*

• În cazul seriilor simple:

$$\sigma^2_z = \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n} = \frac{\sum_{i=1}^{n}\left[(x_i + y_i) - (\bar{x} + \bar{y})\right]^2}{n} = \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x}) + (y_i - \bar{y})\right]^2}{n} =$$

$$= \frac{\sum_{i=1}^{n}\left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + 2(x_i - \bar{x})(y_i - \bar{y})\right]}{n} =$$

$$= \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} + \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} + \frac{2\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} =$$

$$= \sigma^2_x + \sigma^2_y + 2\,cov(x, y)$$

Covarianța este egală cu momentul de ordinul întâi al produsului abaterilor a două variabile față de media lor:

$$cov(x, y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n}$$

Scanned with OKEN Scanner

Covarianţa este, deci, o valoare medie care caracterizează relaţia de dependenţă dintre două variabile.

deci $$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\,cov(x,y) \qquad (7.37)$$

• În cazul seriilor cu frecvenţă:

$$\sigma_z^2 = \frac{\sum_{i=1}^{k}(z_i - \bar{z})^2}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}\left[(x_i + y_i) - (\bar{x} + \bar{y})\right]^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k}\left[(x_i - \bar{x}) + (y_i - \bar{y})\right]^2 n_i}{\sum_{i=1}^{k} n_i} =$$

$$= \frac{\sum_{i=1}^{k}\left[(x_i - \bar{x})^2 n_i + (y_i - \bar{y})^2 n_i + 2(x_i - \bar{x})(y_i - \bar{y})n_i\right]}{\sum_{i=1}^{k} n_i} =$$

$$= \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} + \frac{\sum_{i=1}^{k}(y_i - \bar{y})^2 n_i}{\sum_{i=1}^{k} n_i} + \frac{2\sum_{i=1}^{k}(x_i - \bar{x})(y_i - \bar{y})n_i}{\sum_{i=1}^{k} n_i} =$$

$$= \sigma_x^2 + \sigma_y^2 + 2\,cov(x,y)$$

$$\sigma_z^2 = \sigma_x^2 + \sigma_y^2 + 2\,cov(x,y) \qquad (7.38)$$

Covarianţa pentru două serii cu frecvenţă este dată de expresia:

Scanned with OKEN Scanner

$$cov(x,y)=\frac{\sum_{i=1}^{k}(x_i-\bar{x})(y_i-\bar{y})n_i}{\sum_{i=1}^{k}n_i}$$

- pentru o serie simplă:

$$cov(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n}=\frac{\sum_{i=1}^{n}(x_iy_i-x_i\bar{y}-y_i\bar{x}+\bar{x}\bar{y})}{n}=$$

$$=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n^2}-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n^2}+\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n^2}=$$

$$=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\frac{\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n^2}=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\bar{x}\cdot\bar{y}$$

$$cov(x,y)=\frac{\sum_{i=1}^{n}x_iy_i}{n}-\bar{x}\cdot\bar{y} \qquad (7.39)$$

- pentru o serie cu frecvență:

$$cov(x,y)=\frac{\sum_{i=1}^{k}(x_i-\bar{x})(y_i-\bar{y})n_i}{\sum_{i=1}^{k}n_i}=\frac{\sum_{i=1}^{k}(x_iy_i-x_i\bar{y}-y_i\bar{x}+\bar{x}\bar{y})n_i}{\sum_{i=1}^{k}n_i}=$$

Scanned with OKEN Scanner

$$= \frac{\sum_{i=1}^{k} x_i y_i n_i - \frac{\sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i} - \frac{\sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i} + \frac{\sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i} \cdot \frac{\sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i} \sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} n_i} =$$

$$= \frac{\sum_{i=1}^{k} x_i y_i n_i}{\sum_{i=1}^{k} n_i} - \frac{\sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} y_i n_i}{\sum_{i=1}^{k} n_i} = \frac{\sum_{i=1}^{k} x_i y_i n_i}{\sum_{i=1}^{k} n_i} - \bar{x} \cdot \bar{y}$$

$$cov(x, y) = \frac{\sum_{i=1}^{k} x_i y_i n_i}{\sum_{i=1}^{k} n_i} - \bar{x} \cdot \bar{y} \qquad (7.40)$$

10. *Dacă în suma Z a două variabile X şi Y, prima se înmulţeşte cu o constantă a şi a doua cu b (a şi b putând fi pozitive sau negative), atunci dispersia variabilei sumă va fi:*

$$Z = aX + bY \qquad \text{şi} \qquad z_i = ax_i + by_i$$

$$\sigma_z^2 = \frac{\sum_{i=1}^{n} (z_i - \bar{z})^2}{n} = \frac{\sum_{i=1}^{n} \left[(ax_i + by_i) - (a\bar{x} + b\bar{y}) \right]^2}{n} = \frac{\sum_{i=1}^{n} \left[(ax_i - a\bar{x}) + (by_i - b\bar{y}) \right]^2}{n} =$$

$$= \frac{\sum_{i=1}^{n} (ax_i - a\bar{x})^2 + \sum_{i=1}^{n} (by_i - b\bar{y})^2 + 2\sum_{i=1}^{n} (ax_i - a\bar{x})(by_i - b\bar{y})}{n} =$$

Scanned with OKEN Scanner

$$= \frac{\sum_{i=1}^{n}(ax_i - a\bar{x})^2}{n} + \frac{\sum_{i=1}^{n}(by_i - b\bar{y})^2}{n} + \frac{2\sum_{i=1}^{n}(ax_i - a\bar{x})(by_i - b\bar{y})}{n} =$$

$$= \frac{a\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} + \frac{b\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} + \frac{2ab\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n} =$$

$$= a\sigma_x^2 + b\sigma_y^2 + 2ab\,cov(x,y)$$

deci $$\sigma_z^2 = a\sigma_x^2 + b\sigma_y^2 + 2ab\,cov(x,y) \qquad (7.41)$$

Dacă în această relaţie se substituie $a = 1$ şi $b = -1$ se obţine:

$$Z = aX - bY = X - Y$$

şi $$\sigma_z^2 = \sigma_x^2 - \sigma_y^2 - 2\,cov(x,y) \qquad (7.42)$$

Aplicarea formulei (7.35) o vom ilustra cu ajutorul datelor din tabelul 7.6.

Tabelul 7.6. Calculul simplificat al abaterii medii pătratice

x_i	n_i	$x_i - a$	$\frac{x_i - a}{b}$	$\left(\frac{x_i - a}{b}\right)^2$	$\frac{n_i}{c}$	$\left(\frac{x_i - a}{b}\right)^2 \frac{n_i}{c}$
260	5	-40	-4	16	1	16
270	10	-30	-3	9	2	18
280	20	-20	-2	4	4	16
290	40	-10	-1	1	8	8
300	50	0	0	0	10	0
310	30	10	1	1	6	6
320	25	20	2	4	5	20
330	10	30	3	9	2	18
340	10	40	4	16	2	32
Total	200				40	134

$a = 300 \quad b = 10 \quad c = 5$

Scanned with OKEN Scanner

Substituind în formula (7.35) datele din tabelul 7.6, rezultă:

$$\sigma = \sqrt{\frac{134}{40} \cdot 10^2 - (301-300)^2} = \sqrt{335-1} = 18{,}27 \ piese/zi$$

Constantele a, b şi c se aleg în aşa fel, încât să asigure cea mai mare simplificare a calculelor, astfel:

- a se alege egal cu un termen din centrul şirului de valori ale variabilei X;
- b se consideră egal cu cel mai mare divizor comun al diferenţelor $(x_i - a)$;
- c trebuie să fie egal cu cel mai mare divizor comun al frecvenţelor n_i.

7.2.2. Indicatorii factoriali ai dispersiei

Aşa cum am văzut din prima parte a acestei lucrări, în cadrul colectivităţilor se pot distinge anumite grupe în funcţie de diferite criterii. Aceste grupe pot să reprezinte diferite tipuri din cadrul colectivităţilor, cu o anumită specificitate de dezvoltare sau părţi din colectivitate care se încadrează în anumite intervale de variaţie a caracteristicii. Oricare ar fi cauza împărţirii colectivităţilor în grupe, se pot calcula mai multe categorii de dispersii şi anume:

- o *dispersie generală* calculată la nivelul întregii colectivităţi;
- câte o *dispersie de grupă* pentru fiecare grupă a colectivităţii;
- o *medie* (aritmetică) *a dispersiilor de grupă*;
- o *dispersie a mediilor de grupă faţă de media generală* a colectivităţii.

Pentru a stabili formulele de calcul ale indicatorilor factoriali ai dispersiei, se poate presupune că informaţiile statistice iniţiale asupra unei colectivităţi sunt sistematizate cu ajutorul tabelului 7.7.

Dispersia de grupă, notată cu σ_j^2, se calculează ca o medie a pătratelor abaterilor variantelor variabilei din cadrul unei grupe faţă de media lor, după formula:

$$\sigma_j^2 = \frac{\sum_{i=1}^{k} (x_i - \bar{x}_j)^2 n_{ij}}{\sum_{i=1}^{k} n_{ij}} \qquad i = \overline{1,k} \quad j = \overline{1,p} \qquad (7.43)$$

Scanned with OKEN Scanner

Tabelul 7.7. Distribuția pe grupe și pe niveluri ale caracteristicii

Grupa / X	1	2	...	j	...	p	Total $n_{i.}$
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1p}	$n_{1.}$
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2p}	$n_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{ip}	$n_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{kp}	$n_{k.}$
Total $n_{.j}$	$n_{.1}$	$n_{.2}$	...	$n_{.j}$	...	$n_{.p}$	$n_{..}$
Mediile	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_j$	...	$\bar{x}_p$	$\bar{x}$
Dispersiile	σ_1^2	σ_2^2	...	σ_j^2	...	σ_p^2	σ^2

în care $$\bar{x}_j = \frac{\sum_{i=1}^{k} x_i n_{ij}}{\sum_{i=1}^{k} n_{.j}} \qquad n_{.j} = \sum_{i=1}^{k} n_{ij} \qquad i = \overline{1,k} \quad j = \overline{1,p}$$

Dispersia de grupă măsoară influența factorilor întâmplători, care au acționat asupra caracteristicii unităților statistice dintr-o anumită grupă a colectivității.

Media dispersiilor de grupă, notată cu $\overline{\sigma^2}$, se calculează ca o medie aritmetică a dispersiilor de grupă σ_j^2 ponderate cu volumul grupelor $n_{.j}$ după formula:

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{p} \sigma_j^2 n_{.j}}{\sum_{j=1}^{p} n_{.j}} \qquad j = \overline{1,p} \qquad (7.44)$$

Scanned with OKEN Scanner

în care $n_{.j} = \sum_{i=1}^{k} n_{ij}$ $n_{..} = \sum_{j=1}^{p} n_{.j} = \sum_{i=1}^{k} n_{i.} = \sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij}$

Dacă $n_{.j}$ = constant, atunci $\overline{\sigma^2}$ se calculează ca o medie aritmetică simplă a dispersiilor de grupă. Media dispersiilor de grupă măsoară influenţa medie a factorilor întâmplători care au acţionat în cadrul colectivităţii.

Dispersia dintre grupe, notată cu δ^2, se calculează ca o medie aritmetică a pătratelor abaterilor mediilor de grupă $\bar{x}_j$ faţă de media generală a colectivităţii $\bar{x}$, după formula:

$$\delta^2 = \frac{\sum_{j=1}^{p}(\bar{x}_j - \bar{x})^2 n_{.j}}{\sum_{j=1}^{p} n_{.j}} \qquad (7.45)$$

Dispersia dintre grupe măsoară influenţa factorului de grupare (sistematic) care a determinat separarea grupelor în cadrul colectivităţii.

Dispersia generală, notată cu σ^2, se calculează ca o medie aritmetică a pătratelor abaterilor variantelor unei variabile faţă de media aritmetică a lor, după formula:

$$\sigma^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_{i.}}{\sum_{i=1}^{k} n_{i.}} \qquad (7.46)$$

Dispersia generală reflectă influenţa tuturor factorilor întâmplători şi a factorului de grupare, care au acţionat în cadrul întregii colectivităţi.

Între ultimii trei indicatori ai împrăştierii există o relaţie bine determinată şi anume: *dispersia generală* σ^2 *este egală cu suma dintre media dispersiilor de grupă* $\overline{\sigma^2}$ *şi dispersia dintre grupe* δ^2:

$$\sigma^2 = \overline{\sigma^2} + \delta^2 \qquad (7.47)$$

Scanned with OKEN Scanner

Relaţia (7.47) mai poartă denumirea de *regula de adunare a dispersiilor sau ecuaţia fundamentală a analizei dispersionale.*

7.3. Indicatorii variaţiei pentru o caracteristică alternativă

În vederea ilustrării calculului indicatorilor variaţiei pentru o caracteristică alternativă se consideră tabelul 7.8, în care se prezintă o astfel de distribuţie.

Tabelul 7.8. Distribuţia unei caracteristici alternative

Caracteristica x_i	Numărul unităţilor n_i	Frecvenţa relativă f_i	$x_i f_i$
1	n_1	p	p
0	n_2	q	0
Total	n	1	p

După cum am văzut, valoarea medie a unei caracteristici alternative este egală cu greutatea specifică a unităţilor statistice care poartă caracteristica:

$$\bar{x}_p = \frac{\sum_{i=1}^{2} x_i n_i}{\sum_{i=1}^{2} n_i} = \frac{n_1}{n} \quad \text{sau} \quad \bar{x}_p = \frac{\sum_{i=1}^{2} x_i f_i}{\sum_{i=1}^{2} f_i} = p \qquad i = 1,2$$

Aplicând formula generală a dispersiei şi în cazul caracteristicii alternative, se obţine:

$$\sigma_p^2 = \frac{\sum_{i=1}^{2} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{2} f_i} = \frac{(0-p)^2 q + (1-p)^2 p}{p+q} = \frac{p^2 q + q^2 p}{p+q} = pq$$

Scanned with OKEN Scanner

deci $$\sigma_p^2 = pq \qquad (7.48)$$

sau $$\sigma_p^2 = p(1-p) \qquad (7.49)$$

Din formula (7.49) rezultă că dispersia unei caracteristici alternative este egală cu produsul între ponderea unităţilor care posedă nivelul x_1 al caracteristicii şi ponderea celor care posedă nivelul x_2 al acesteia.

Abaterea medie pătratică a caracteristicii alternative va fi prin definiţie:

$$\sigma_p = \sqrt{pq} \quad \text{sau} \quad \sigma_p = \sqrt{p(1-p)} \qquad (7.50)$$

În particular, dacă numărul unităţilor care posedă nivelul x_1 al caracteristicii este egal cu numărul unităţilor care posedă celălalt nivel al ei x_2, adică $p = q = 50\%$ atunci dispersia va fi egală cu:

$$\sigma_p^2 = 0{,}5 \cdot 0{,}5 = 0{,}25$$

$\sigma_p^2 = 0{,}25$ *este valoarea maximă pe care o poate lua dispersia unei caracteristici alternative.*

Indicatorii variaţiei, calculaţi pentru caracteristici alternative, îşi găsesc o largă aplicare în studiul statistic al calităţii produselor, în cercetările prin sondaj din economie, biologie, psihologie, pedagogie şi, în general, în toate domeniile în care însuşirile fenomenelor pot fi privite drept caracteristici alternative, binare. Din punct de vedere teoretic, orice caracteristică nealternativă se poate transforma într-o caracteristică alternativă, prin determinarea abaterilor nivelurilor individuale ale acesteia faţă de nivelul ei mediu sau faţă de o altă valoare reprezentativă a seriei şi prin împărţirea abaterilor, astfel calculate, în două grupe după semnul lor (abateri pozitive şi abateri negative).

7.4. Concentrarea

Concentrarea variantelor unei variabile în jurul unei valori a tendinţei centrale este un fenomen opus dispersării. Studiul concentrării permite analiza structurală a colectivităţilor a căror caracteristică este o mărime pozitivă şi cumulativă. Astfel se poate cerceta distribuţia întreprinderilor după mărimea

Scanned with OKEN Scanner

veniturilor, după cifra afacerilor, după suprafaţa cultivată, distribuţia populaţiei pe localităţi, pe zone administrative etc.

Studiul concentrării urmăreşte să caracterizeze gradul de apropiere (respectiv de depărtare) a unei distribuţii concrete, observate faţă de distribuţia egalitară corespunzătoare a caracteristicii luate în considerare.

În general, o *repartiţie egalitară* a volumului unei caracteristici între unităţile statistice ale unei colectivităţi se prezintă în aşa fel încât $n\%$ dintre unităţile statistice posedă $n\%$ din volumul caracteristicii, oricare ar fi valoarea lui n. Fie o serie statistică:

$$X: \begin{pmatrix} x_1, & x_2, & \ldots, & x_i, & \ldots, & x_k \\ n_1, & n_2, & \ldots, & n_i, & \ldots, & n_k \end{pmatrix}$$

Frecvenţele cumulate:

$$N_i = \sum_{i=1}^{k} n_i \qquad \forall\, i = \overline{1,k}$$

Produsele cumulate:

$$L_i = L(x) = \sum_{i=1}^{k} x_i n_i$$

$$p_i(\%) = \frac{\sum_{i=1}^{i} x_i n_i}{\sum_{i=1}^{k} x_i n_i} \qquad \text{iar} \qquad q_i = \frac{N_i}{\sum_{i=1}^{k} n_i}$$

Reprezentând într-un sistem de axe rectangulare (fig.7.3) cele două variabile p_i şi q_i pe ordonată şi respectiv pe abscisă, bisectoarea axelor de coordonate OA va reprezenta distribuţia egalitară a volumului caracteristicii studiate, iar funcţia $f(q_i, p_i)$ ia forma unei curbe ODA cu atât mai depărtată de diagonala pătratului cu cât distribuţia studiată va fi mai concentrată, mai inegalitară. Această curbă poartă denumirea de *curbă de concentrare* sau *curba Lorenz-Gini*, după numele celor doi statisticieni (primul, americanul *Lorenz*, care în anul 1905 a publicat-o în articolul " *Method of measuring the concentration of wealth*" în *Journal of the American Statistical Association*, iunie, 1905; iar al doilea, în comunicarea " *Indici di concentrazione et di dependenza*", apărută în anul 1910).

Scanned with OKEN Scanner

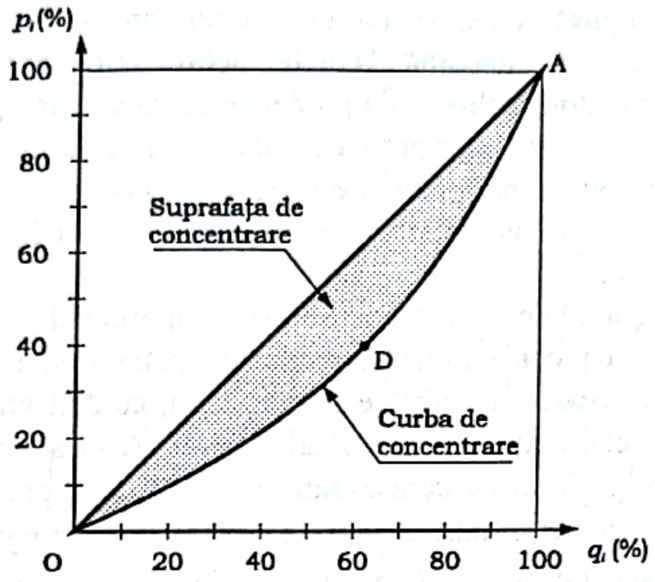


Figura 7.3. Curba de concentrare şi suprafaţa de concentrare

Relaţia dintre mediană, medială şi concentrarea unei variabile se poate ilustra şi pe cale grafică (fig.7.4).

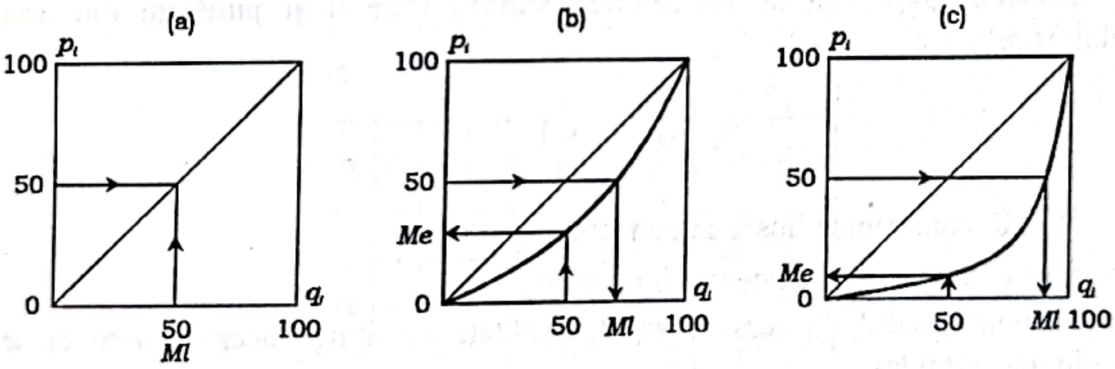


Figura 7.4. Gradul de concetrare

Din figura 7.4.a se observă că, în cazul în care curba de concentrare se confundă cu diagonala pătratului, nu există nici un fel de concentrare, iar abaterea dintre mediană şi medială este nulă. Absenţa concentrării înseamnă, în cazul unei distribuţii a muncitorilor unei întreprinderi după producţia zilnică obţinută, că $m\,\%$ dintre muncitori realizează $m\,\%$ din producţia totală. Există deci o concordanţă perfectă între numărul de muncitori şi producţia obţinută. Acest tip de distribuţie

Scanned with OKEN Scanner

poartă denumirea de *distribuţie egalitară* sau de *echirepartiţie*, căreia îi corespunde o concentrare nulă.

În situaţia opusă, când curba de concentrare se confundă cu laturile pătratului, concentrarea este maximă. Teoretic, acest caz extrem se poate interpreta astfel: 99% dintre muncitori realizează o producţie egală cu zero şi un singur procent dintre muncitori realizează întreaga producţie 100%.

Între aceste două situaţii extreme se situează cazurile reale de concentrare. Figura 7.4.b reprezintă o repartiţie cu o slabă concentrare, iar figura 7.4.c una cu un grad mare de concentrare.

În general, gradul de concentrare variază proporţional cu mărimea suprafeţei cuprinse între curba de concentrare şi diagonala pătratului, numită *suprafaţă de concentrare*. Cu cât această suprafaţă este mai mică, cu atât gradul de concentrare este mai redus şi invers. Statisticianul italian *Corrado Gini* a formulat pentru prima dată un indicator al gradului de concentrare în raport cu suprafaţa de concentrare, numit *indice de concentrare*. Indicele de concentrare i_c este egal cu raportul dintre mărimea suprafeţei de concentrare C şi jumătate din aria pătratului $S/2$:

$$i_c = \frac{suprafata\ de\ concentrare}{\frac{aria\ pătratului}{2}} \qquad (7.44)$$

Deoarece suprafaţa de concentrare variază între 0 şi jumătate din aria pătratului, rezultă că:

$$0 \le i_c \le 1$$

$i_c = 0$ corespunde lipsei de concentrare;

$i_c = 1$ corespunde concentrării maxime.

Desigur, gradul de concentrare în realitate nu atinge aceste limite ci se situează în interiorul lor.

Pentru exemplificare, considerăm distribuţia întreprinderilor din *România* după numărul muncitorilor în anul 1995 (tabelul 7.9 coloanele 1 şi 2).

Deoarece caracteristica este dată sub formă de intervale neegale, pentru determinarea numărului real al muncitorilor $x_i n_i$ care aparţin întreprinderilor dintr-un anumit interval, s-au utilizat frecvenţele relative ale acestor intervale în funcţie de numărul total al muncitorilor 3198081, în locul produselor dintre mijloacele intervalelor şi numărul întreprinderilor, care diferă în mare măsură de realitate.

Scanned with OKEN Scanner

Tabelul 7.9. Întreprinderile industriale după numărul salariaţilor în anul 1995

Grupe de întreprinderi după numărul de salariaţi	Numărul întreprinderilor		Numărul salariaţilor		$N_i = \sum_i n_i$	$q_i = \frac{N_i}{\sum_{i=1}^{i} n_i}$ %	$L_i = \sum_{1i=1}^{i} x_i n_i$	$p_i = \frac{L_i}{\sum_{i=1}^{n} x_i n_i}$ %
	Total n_i	%	Total $x_i n_i$	%				
- 100	32 279	92,1	177 803	6,8	32 279	92,0	177 803	6,8
101 - 200	677	1,9	94 131	3,6	32 956	94,0	271 934	10,4
201 - 300	359	1,0	86 287	3,3	33 315	95,0	358 221	13,7
301 - 400	224	0,6	75 827	2,9	33 539	95,6	434 048	16,6
401 - 500	222	0,6	96 746	3,7	33 761	96,3	530 794	20,3
501 - 1000	583	1,7	405 286	15,5	34 344	97,9	936 080	35,8
1001 - 2000	435	1,2	601 392	23,0	34 779	99,2	1 537 472	58,8
2001 - 3000	158	0,5	371 294	14,2	34 937	99,6	1 908 766	73,0
3001 - 5000	93	0,3	339 917	13,0	35 030	99,9	2 248 683	86,0
5001 -	40	0,1	366 063	14,0	35 070	100,0	2 614 746	100,0
Total	35 070	100	2 614 746	100	-	-	-	-

Sursa: *Anuarul statistic al României 1996*, CNS, Bucureşti, p. 485.

Curba şi suprafaţa de concentrare se prezintă în figura 7.5.

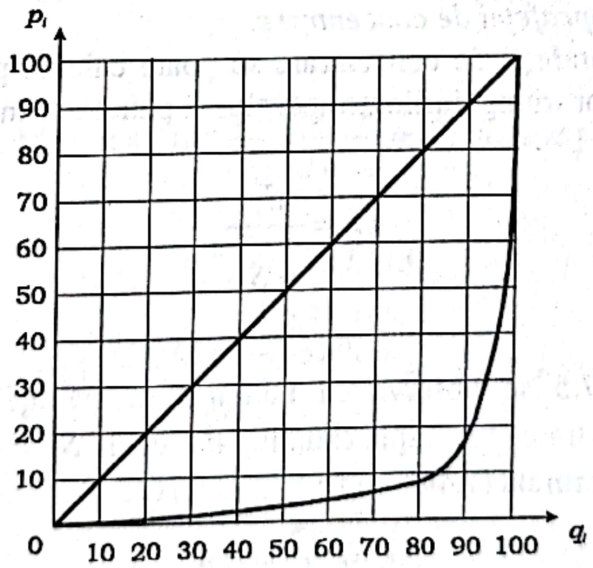


Figura 7.5. Concentrarea întreprinderilor industriale din *România* după numărul salariaţilor în anul 1995

Scanned with OKEN Scanner

Există o corespondenţă între concentrare şi curba frecvenţelor cumulate, pe baza căreia se pot evidenţia cei doi indicatori ai tendinţei centrale (mediana şi mediala) cu ajutorul cărora se poate aprecia şi gradul de concentrare a unei distribuţii (fig. 7.6).

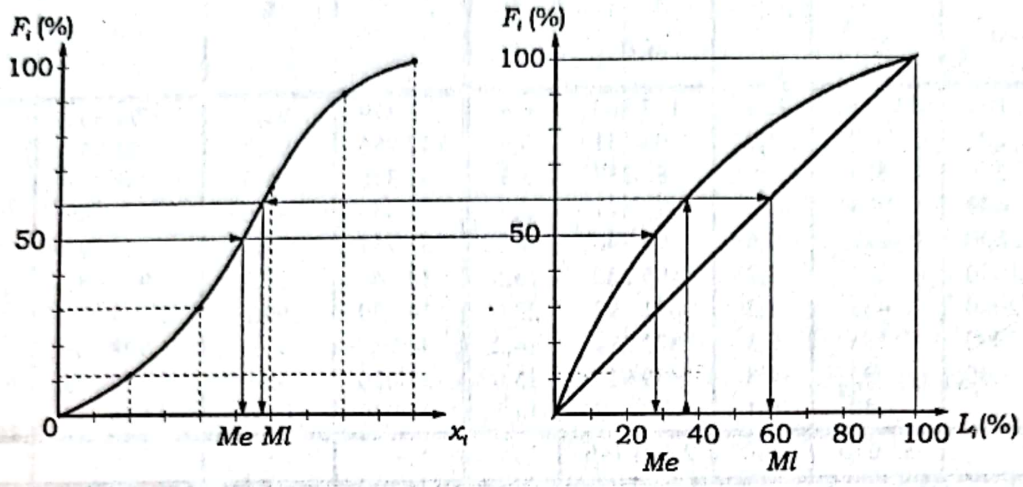


Figura 7.6. Curba de concentrare şi poligonul frecvenţelor cumulate

Indicele de concentrare se poate determina în mai multe feluri:

1) *Pe baza suprafeţei de concentrare.*

Mărimea suprafeţei de concentrare se poate calcula pe cale grafică prin însumarea suprafeţelor triunghiurilor, trapezelor şi pătratelor înscrise în interiorul ei cu ajutorul formulei:

$$I_G = \frac{S_c}{\frac{1}{2} S_p} \qquad (7.45)$$

Din figura 7.5 se observă că numărul de pătrăţele mici cuprinse în suprafaţa de concentrare este aproximativ 45, deci $S_c = 45$, iar indicele de concentrare, dat de formula (7.44), va fi:

$$i_c = \frac{45}{50} = 0{,}9$$

Aceasta înseamnă că industria *României* din anul 1995 se caracterizează printr-un foarte ridicat grad de concentrare spre întreprinderile mici şi foarte mici, cu toate consecinţele economice nefavorabile care decurg de aici.

Scanned with OKEN Scanner

2) Gradul de concentrare se poate aprecia şi cu ajutorul *diferenţei dintre mediană şi medială*, notată cu $\Delta_{Me,Ml}$:

$$\Delta_{Me,Ml} = Ml - Me \geq 0 \qquad (7.46)$$

Pentru ca acest indicator să poată fi utilizat în studiul comparativ al gradului de concentrare se determină expresia relativă a diferenţei absolute $\Delta_{Me,Ml}$ prin împărţirea acesteia la Me şi se obţine:

$$k_{\Delta} = \frac{\Delta_{Me,Ml}}{Me} = \frac{Ml - Me}{Me} \geq 0 \qquad (7.47)$$

Dacă $k_{\Delta} = 0$ înseamnă că $Me = Ml$ şi distribuţia este egalitară, iar dacă $k_{\Delta} > 0$ distribuţia studiată prezintă un anumit grad de concentrare care este cu atât mai mare cu cât k_{Δ} este mai mare şi invers.

Calculând k_{Δ} exemplului din tabelul 7.9 rezultă:

$$U^{me} = \frac{\sum_i n_i + 1}{2} = \frac{35070}{2} = 17535$$

$$Me = 1 + 100 \cdot \frac{17535 - 0}{35070} = 51 \ salariaţi$$

$$U^{ml} = 1307373$$

$$Ml = 500 + 500 \cdot \frac{1307373 - 936080}{601392} = 808{,}7 \approx 809 \ salariaţi$$

Aplicând formula (7.47) rezultă:

$$k_{\Delta} = \frac{Ml - Me}{Me} = \frac{809 - 51}{51} = 14{,}86$$

$k_{\Delta} = 14{,}86$ arată de asemenea că industria *României* în anul 1995 prezenta un înalt grad de concentrare spre întreprinderile mici şi foarte mici.

Scanned with OKEN Scanner

7.4.1. Proprietăţile curbei de concentrare

Considerând variabila X şi notând: $q = f(p)$

unde

$$q = \frac{\textit{suma variantelor mai mici decât } x}{\textit{suma totală a variantelor caracteristicii}}$$

sau

$$q = \frac{\textit{media variantelor inferioare lui } x}{\textit{media tuturor variantelor caracteristicii}}$$

Nivelul mediu al caracteristicii pe total distribuţie se poate determina astfel:

- în cazul X discret:

$$\bar{x} = \frac{\sum_{1}^{k} x_i n_i}{\sum_{1}^{k} n_i} = \sum_{1}^{k} x_i f_i$$

- în cazul X continuu:

$$\bar{x} = \int_{-\infty}^{+\infty} xf(x)dx = \int_{x_{min}}^{x_{max}} xf(x)dx$$

unde $f(x)$ este curba frecvenţelor.

Nivelul mediu pentru variantele mai mici decât x se va putea scrie:

$$\bar{x}' = \int_{x_{min}}^{x} vf(v)dv$$

Scanned with OKEN Scanner

deci $$q = \frac{\bar{x}'}{\bar{x}} = \frac{\int_{x_{min}}^{x} vf(v)dv}{\int_{x_{min}}^{x_{max}} xf(x)dx}$$

Coordonatele punctelor curbei de concentrare vor fi deci:

$$\begin{cases} p = \int_{x_{min}}^{x} f(v)dv \\ q = \frac{\int_{x_{min}}^{x} vf(v)dv}{\bar{x}} \end{cases}$$

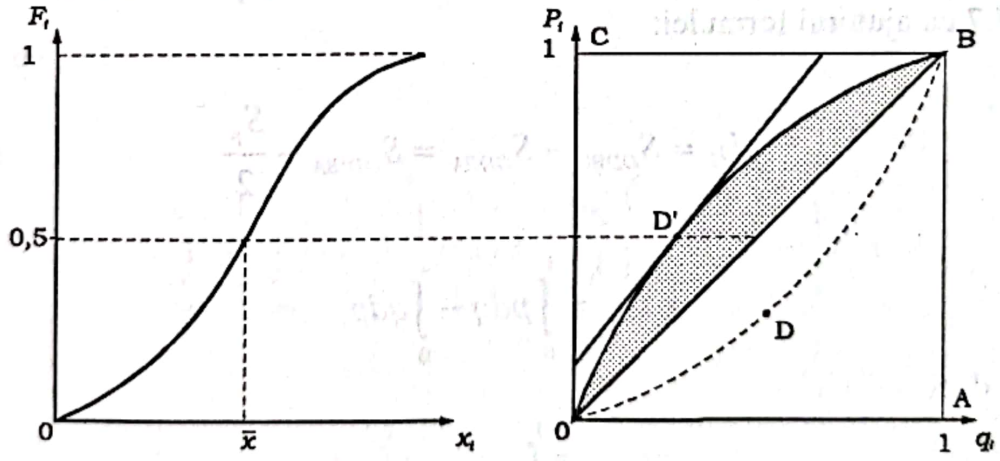


Figura 7.7. Curba frecvenţelor cumulate şi tangenta curbei de concentrare

Scanned with OKEN Scanner

iar panta curbei de concentrare este derivata în raport cu p a lui $q = f(p)$:

$$\frac{dq}{dp} = \frac{dq}{dx} \cdot \frac{dx}{dp} = \frac{dq}{dx} \cdot \frac{1}{\frac{dp}{dx}}$$

$$\frac{dq}{dp} = \frac{1}{\bar{x}} \cdot x \cdot f(x) \cdot \frac{1}{f(x)} = \frac{x}{\bar{x}}$$

Analizând panta curbei de concentrare, rezultă că, pentru $x = x_{min} = 0$, tangenta curbei de concentrare este chiar axa ordonatelor, iar pentru $x = x_{max} = \infty$, tangenta la curba de concentrare este o paralelă cu axa absciselor şi anume CB; în timp ce, pentru $x = \bar{x}$, tangenta la curba de concentrare este o paralelă cu bisectoarea axelor de coordonate.

7.4.2. Indicele de concentrare *Gini*

Algebric, indicele de concentrare al lui *C.Gini* se poate determina pe baza figurii 7.7 cu ajutorul formulei:

$$I_G = S_{ODBC} - S_{ODBA} = S_{OD'BA} - \frac{S_p}{2} \qquad (7.48)$$

$$I_G = \int_0^1 pdq - \int_0^1 qdp$$

deci

$$\bar{x} I_G = \bar{x} \int_0^1 pdq - \bar{x} \int_0^1 qdp$$

dar $\frac{dq}{dp} = \frac{x}{\bar{x}} \Rightarrow dp = \frac{\bar{x}}{x} \cdot dq$ şi $dq = \frac{x}{\bar{x}} \cdot dp$

Scanned with OKEN Scanner

rezultă $$\bar{x}I_G = \bar{x}\int_0^1 p\frac{x}{\bar{x}}dp - \bar{x}\int_0^1 qdp$$

când $p = 0$ $x = x_{min}$

$p = 1$ $x = x_{max}$

Dar q este o funcţie de x ,deoarece reprezintă ponderea sumei variantelor caracteristicii X mai mici decât x.

Se poate scrie că:

$$\bar{x}I_G = \bar{x}\int_{x_{min}}^{x_{max}} p\frac{x}{\bar{x}}f(x)dx - \bar{x}\int_{x_{min}}^{x_{max}} qf(x)dx = \int_{x_{min}}^{x_{max}} pxf(x)dx - \int_{x_{min}}^{x_{max}} \bar{x}qf(x)dx$$

dar $p = \int_{x_{min}}^{x} f(v)dv$ şi $\bar{x}q = \int_{x_{min}}^{x} vf(v)dv$

de unde

$$\bar{x}I_G = \int_{x_{min}}^{x_{max}} x\left(\int_{x_{min}}^{x} f(v)dv\right)f(x)dx - \int_{x_{min}}^{x_{max}}\left(\int_{x_{min}}^{x} vf(v)dv\right)f(x)dx$$

Este evident că $$\int_{x_{min}}^{x_{max}}\int_{x_{min}}^{x_{max}}(x-v)f(x)f(v)dxdv = 0$$

Scanned with OKEN Scanner

rezultă

$$\bar{x}I_G = \frac{1}{2}\int_{x_{min}}^{x_{max}}\int_{x_{min}}^{x}(x-v)f(x)f(v)dxdv + \frac{1}{2}\int_{x_{min}}^{x_{max}}\int_{x_{min}}^{x_{max}}(v-x)f(x)f(v)dxdv$$

sau $$\bar{x}I_G = \frac{1}{2}\int_{x_{min}}^{x_{max}}\int_{x_{min}}^{x_{max}}|x-v|f(x)f(v)dxdv$$

Deci, în cazul unei variabile continue, formula de calcul al indicelui de concentrare va fi:

$$I_G = \frac{\int_{x_{min}}^{x_{max}}\int_{x_{min}}^{x_{max}}|x-v|f(x)f(v)dxdv}{2\bar{x}} \qquad (7.49)$$

Iar în cazul unei variabile discrete, formula de calcul va lua expresia:

$$I_G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j|n_i n_j}{2n(n-1)\bar{x}} \qquad (7.50)$$

Calculul sumei duble de la numărător se poate efectua sistematizat în mai multe etape prin compararea fiecărei variante x_i a variabilei cu fiecare variantă x_j a ei .

Într-o serie, formată din n variante ale variabilei X, cele n^2 valori absolute ale diferenţelor $|x_i - x_j|$ se pot scrie astfel:

Scanned with OKEN Scanner

Tabelul 7.10. Calculul diferenţelor $|x_i - x_j|$

x_j \ x_i	x_1	x_2	...	x_i	...	x_n
x_1	0	$\|x_2 - x_1\|$	...	$\|x_i - x_1\|$	...	$\|x_n - x_1\|$
x_2	$\|x_1 - x_2\|$	0	...	$\|x_i - x_2\|$	...	$\|x_n - x_2\|$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_j	$\|x_1 - x_j\|$	$\|x_2 - x_j\|$	...	0	...	$\|x_n - x_j\|$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
x_n	$\|x_1 - x_n\|$	$\|x_2 - x_n\|$	...	$\|x_i - x_n\|$	...	0

Tabelul 7.11. Calculul produselor $n_i \cdot n_j$

n_j \ n_i	n_1	n_2	...	n_i	...	n_n
n_1	n_1^2	$n_2 n_1$	...	$n_i n_1$	...	$n_n n_1$
n_2	$n_1 n_2$	n_2^2	...	$n_i n_2$	...	$n_n n_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n_j	$n_1 n_j$	$n_2 n_j$	...	$n_i n_j$	...	$n_n n_j$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n_n	$n_1 n_n$	$n_2 n_n$	...	$n_i n_n$	...	n_n^2

Deoarece $|x_i - x_i| = 0, \quad \forall\, i = \overline{1,n}$, rezultă că numărul diferenţelor diferite de zero este egal cu $n(n-1)$, fiecare variantă x_i a caracteristicii X se compară cu toate celelalte variante ale acestei variabile.

Scanned with OKEN Scanner

Tabelul 7.12. Calculul expresiei $|x_i - x_j| n_i n_j$

j \ i	$\lvert x_1 - x_j\rvert n_1 n_j$	...	$\lvert x_i - x_j\rvert n_i n_j$	...	$\lvert x_n - x_j\rvert n_n n_j$	$\sum_i \lvert x_i - x_j\rvert n_i n_j$
$\lvert x_i - x_1\rvert n_i n_1$	0	...	$\lvert x_i - x_1\rvert n_i n_1$	...	$\lvert x_n - x_1\rvert n_n n_1$	$\sum_i \lvert x_i - x_1\rvert n_i n_1$
$\lvert x_i - x_2\rvert n_i n_2$	$\lvert x_1 - x_2\rvert n_1 n_2$	...	$\lvert x_i - x_2\rvert n_i n_2$	...	$\lvert x_n - x_2\rvert n_n n_2$	$\sum_i \lvert x_i - x_2\rvert n_i n_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\lvert x_i - x_j\rvert n_i n_j$	$\lvert x_1 - x_j\rvert n_1 n_j$	...	0	...	$\lvert x_n - x_j\rvert n_n n_j$	$\sum_i \lvert x_i - x_j\rvert n_i n_j$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$\lvert x_i - x_n\rvert n_i n_n$	$\lvert x_1 - x_n\rvert n_1 n_n$	...	$\lvert x_i - x_n\rvert n_i n_n$	...	0	$\sum_i \lvert x_i - x_n\rvert n_i n_n$
$\sum_j \lvert x_i - x_j\rvert n_i n_j$	$\sum_j \lvert x_1 - x_j\rvert n_1 n_j$	...	$\sum_j \lvert x_i - x_j\rvert n_i n_j$	...	$\sum_j \lvert x_n - x_j\rvert n_n n_j$	$\sum_i \sum_j \lvert x_i - x_j\rvert n_i n_j$

Pentru exemplificarea calculului lui I_G, considerăm din nou seria utilizată pentru construirea curbei de concentrare, luând drept variante x_i ale variabilei mijloacele intervalului de grupare.

Aplicarea formulei (7.50) necesită efectuarea unor calcule auxiliare, care se pot sistematiza în tabelele 7.13, 7.14 şi 7.15.

Tabelul 7.13. Calculul diferenţelor $|x_i - x_j|$

x_j \ x_i	50	150	250	350	450	750	1500	2500	4000	6000
50	0	100	200	300	400	700	1450	2450	3950	5950
150	100	0	100	200	300	600	1350	2350	3850	5850
250	200	100	0	100	200	500	1250	2250	3750	5750
350	300	200	100	0	100	400	1150	2150	3650	5650
450	400	300	200	100	0	300	1050	2050	3550	5550
750	700	600	500	400	300	0	750	1750	3250	5250
1500	1450	1350	1250	1150	1050	750	0	1000	2500	4500
2500	2450	2350	2250	2150	2050	1750	1000	0	1500	3500
4000	3950	3850	3750	3640	3550	3250	2500	1500	0	2000
6000	5950	5850	5750	5650	5550	5250	4500	3500	2000	0

Scanned with OKEN Scanner

Tabelul 7.14. Calculul produselor $n_i n_j$

n_j \ n_i	32279	677	359	224	222
0	1	2	3	4	5
32279	1041933841	21852883	11588161	7230496	7165938
677	21852883	458329	243043	151648	150294
359	11588161	243043	12888	80416	79698
224	7230496	151648	80416	50176	49728
222	7165938	150294	79698	49728	49284
583	18818657	394691	209297	130592	129426
435	14041365	294495	156165	97440	96570
158	5100082	106966	56722	35392	35076
93	3001947	62961	33387	20832	20646
40	1291160	27080	14360	8960	8880

Tabelul 7.14 continuare

583	435	158	93	40
6	7	8	9	10
18818657	14041365	5100082	3001947	1291160
394691	294495	106966	62961	27080
209297	156165	56722	33387	14360
130592	97440	35392	20832	8960
129426	96570	35076	20646	8880
339889	253605	92114	54218	23320
253605	189225	68730	40455	17400
92114	68730	24964	14694	6320
54219	40455	14694	8649	3720
23320	17400	6320	3720	1600

Scanned with OKEN Scanner

Tabelul 7.15. Calculul produselor $|x_i - x_j| n_i n_j$

i / j	$\|x_1 - x_j\| n_1 n_j$	$\|x_2 - x_j\| n_2 n_j$	$\|x_3 - x_j\| n_3 n_j$	$\|x_4 - x_j\| n_4 n_j$	$\|x_5 - x_j\| n_5 n_j$
1	2	3	4	5	6
$\|x_1 - x_i\| n_1 n_i$	0	2 185 288 300	2 317 632 00	2 169 148800	2 866 375 200
$\|x_2 - x_i\| n_2 n_i$	2 185 288 300	0	24 304 300	30 329 600	45 088 200
$\|x_3 - x_i\| n_3 n_i$	2 317 632 200	24 340 300	0	8 041 600	15 939 600
$\|x_4 - x_i\| n_4 n_i$	2 169 148 800	30 329 600	8 041 600	0	4 972 800
$\|x_5 - x_i\| n_5 n_i$	2 866 375 200	45 088 200	15 939 600	4 972 800	0
$\|x_6 - x_i\| n_6 n_i$	13 173 059 900	236 814 600	104 648 500	52 236 800	38 827 800
$\|x_7 - x_i\| n_7 n_i$	20 359 979 250	397 568 250	195 206 250	112 056 000	101 398 500
$\|x_8 - x_i\| n_8 n_i$	12 495 200 900	251 370 100	127 624 500	76 092 800	71 905 800
$\|x_9 - x_i\| n_9 n_i$	11 857 690 650	242 399 850	125 201 250	76 036 800	73 293 300
$\|x_{10} - x_i\| n_{10} n_i$	7 682 402 000	158 418 000	82 570 000	50 624 000	49 284 000
Total	75 106 777 200	3 571 581 200	3 001 168200	2 579 539200	3 267 085 200

continuare tabel 7.15

$\|x_6 - x_j\| n_6 n_j$	$\|x_7 - x_j\| n_7 n_j$	$\|x_8 - x_j\| n_8 n_j$	$\|x_9 - x_j\| n_9 n_j$	$\|x_{10} - x_j\| n_{10} n_j$	Total
7	8	9	10	11	12
13 173 059 900	20 359 979 250	12 495 200 900	11 857 690 650	7 682 402 000	75 106 777 200
236 814 600	397 568 250	251 370 100	242 399 850	158 418 000	3 571 581 200
104 648 500	195 206 250	127 624 500	125 201 250	82 570 000	3 001 168 200
52 236 800	112 056 000	76 092 800	76 036 800	50 624 000	2 579 539 200
38 827 800	101 398 500	71 905 800	73 293 300	49 284 000	3 267 085 200
0	190 203 750	161 199 500	176 211 750	122 430 000	14 255 632 600
190 203 750	0	68 730 0000	101 137 500	78 300 000	21 604 579 500
161 199 500	68 730 000	0	22 041 000	22 120000	13 296 284 600
176 211 750	101 137 500	22 041 000	0	7 440 000	12 681 452 100
122 430 000	78 300 000	22 120 000	7 440 000	0	82 253 588 000
14 255 632 000	21 604 579 500	13 296 284 600	12 681 452 100	8 253 588 000	157 617 687 800

Scanned with OKEN Scanner

Substituind în formula (7.50) datele din tabelul 7.12., rezultă:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i n_i}{\sum_{i=1}^{n} n_i} = \frac{2614746}{35070} = 74{,}6 \ \textit{salariaţi/întreprindere}$$

$$I_G = \frac{157\,617\,687\,800}{2 \cdot 35070 \cdot (35070 - 1) \cdot 74{,}6} = 0{,}86$$

$I_G = 0{,}86$ arată că industria românească din anul 1995 se caracterizează printr-un foarte ridicat grad de concentrare îndreptat spre întreprinderile mici. De altfel, se observă din tabelul 7.9 că peste 92% din întreprinderi au numărul de salariaţi mai mic de 100. Deşi foarte numeroase, întreprinderile mici concentrează doar 6,8% din salariaţii întregii industrii.

Indicatorul propus de *C.Gini* pentru a măsura gradul de concentrare a variantelor unei distribuţii este mai mult un indicator de măsurare a inegalităţii distribuţiei, adică a abaterii distribuţiei studiate faţă de o distribuţie perfect egalitară.

Coeficientul lui *Gini* nu ia în considerare mişcările de concentrare. Astfel, dacă un anumit număr de întreprinderi de aceeaşi mărime fuzionează pentru a forma un număr mai mic de întreprinderi de aceeaşi mărime (evident mai mare decât în primul caz) se înţelege că are loc un proces de concentrare, dar indicele lui *Gini* nu se modifică.

În cazuri particulare indicatorul lui *Gini* poate chiar denatura realitatea.

De exemplu: să presupunem că la un moment dat există într-o zonă 100 de întreprinderi cu numărul de salariaţi cuprins între 10 şi 50. După o anumită perioadă de timp, are loc un proces de fuzionare prin care mărimea întreprinderilor se apropie într-o mai mare măsură faţă de mărimea medie. În acest caz, coeficientul lui *Gini* va indica o diminuare a gradului de concentrare evident contrar realităţii. În plus, două feluri de inegalităţi pot fi exprimate prin acelaşi indicator *Gini*, deoarece suprafaţa de concentrare rămâne neschimbată în ambele cazuri (fig.7.8).

Un indicator corespunzător al concentrării activităţii într-un domeniu trebuie să ia în calcul atât numărul de întreprinderi, cât şi inegalitatea mărimii lor dată de importanţa relativă a cifrei lor de afaceri, de volumul producţiei sau de numărul de salariaţi.

Scanned with OKEN Scanner

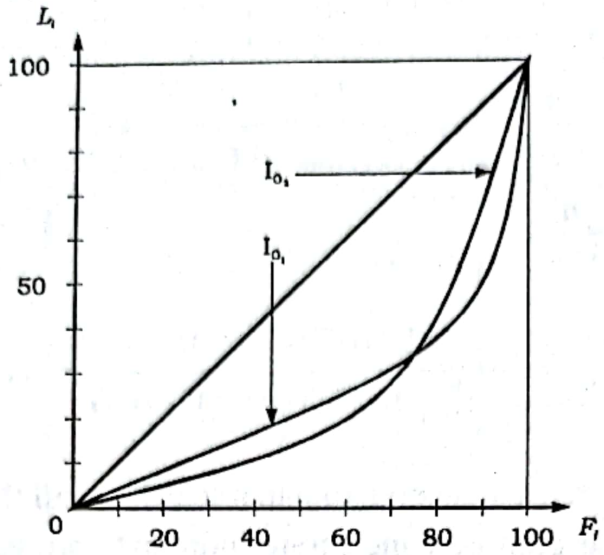


Figura 7.8: Curbe de concentrare diferite care delimitează aceeaşi suprafaţă de concentrare

Astfel, dacă o anumită ramură a industriei cuprinde n întreprinderi ordonate în ordine descrescătoare a mărimii producţiei lor q_i, atunci producţia totală a acelei ramuri va fi:

$$Q = \sum_{i=1}^{n} q_i$$

Fie g_i ponderea fiecarei înteprinderi faţă de total:

$$g_i = \frac{q_i}{\sum_{1}^{n} q_i}$$

dacă toate întreprinderile ar fi de aceeaşi mărime ar rezulta:

$$g_i = \frac{q_i}{\sum_{1}^{n} q_i} = \frac{q}{nq} = \frac{1}{n}$$

Scanned with OKEN Scanner

În general există o relaţie inversă între numărul de întreprinderi şi importanţa lor relativă şi de aceea mărimea inversă a numărului de firme $\frac{1}{n}$ nu poate să se constituie într-un indicator corespunzător al concentrării, deoarece acesta nu ia în calcul inegalităţile de mărime.

Scanned with OKEN Scanner

8. REPARTIŢIILE UNOR VARIABILE ALEATOARE

Alături de distribuţiile empirice, în statistică se acordă o atenţie deosebită distribuţiilor teoretice.

Distribuţia unei variabile aleatoare a unei colectivităţi este determinată, dacă se cunoaşte mulţimea variantelor x_i ale ei şi funcţia de probabilitate $f(x_i)$ sau densitatea de probabilitate $\varphi(x_i)$, după cum distribuţia este continuă sau discretă. În practica statistică, $f(x_i)$ sau $\varphi(x_i)$ se pot determina fie pe baza unor distribuţii teoretice, fie pe baza unei distribuţii empirice.

Prima distribuţie teoretică a fost introdusă în ştiinţă de către *Jacques Bernoulli* (1654-1705), în anul 1700 şi poartă denumirea de *distribuţie binomială*, urmează apoi *distribuţia normală* datorată lui *Abraham De Moivre* (1667-1754) şi *distribuţia Poisson*, descoperită de către *Simion Denis Poisson* (1781-1840) în anul 1837.

Atât din punct de vedere istoric, cât şi al importanţei lor în statistică aceste distribuţii ocupă un loc aparte şi sunt denumite în mod obişnuit distribuţii clasice. Pe lângă distribuţiile clasice şi pe baza lor, au fost introduse în statistică multe alte distribuţii teoretice particulare.

8.1. Repartiţia binomială

Considerând schema probabilistică clasică a urnei lui *Bernoulli* cu două stări, se poate calcula probabilitatea ca în n extracţii independente să se verifice de k ori apariţia unei bile de o anumită culoare, după următoarea formulă:

$$P_n(k) = C_n^k p^k q^{n-k} \qquad k = \overline{0,n} \tag{8.1}$$

unde $$p + q = 1$$

Se observă că $P_n(k)$ reprezintă termenul general al dezvoltării binomului lui *Newton*:

Scanned with OKEN Scanner

$$(p+q)^n = C_n^n p^n + C_n^{n-1} p^{n-1} q + \cdots + C_n^k p^k q^{n-k} + \cdots + C_n^0 q^n \qquad (8.2)$$

Primul termen al dezvoltării binomului din (8.2) reprezintă probabilitatea ca o bilă de o anumită culoare să apară în toate cele n extrageri, al doilea termen exprimă probabilitatea ca bila respectivă să apară în $n-1$ extrageri şi aşa mai departe, ultimul termen indicând faptul că bila respectivă nu apare în nici o extragere, deoarece în toate apare cealaltă bilă.

Astfel, dacă facem patru extracţii, o bilă de o anumită culoare (în urnă existând bile de două culori) poate să apară de patru ori, de trei ori, de două ori, o singură dată sau niciodată, iar probabilităţile corespunzătoare vor fi:

$$C_4^4 p^4 q^0 = \frac{4!}{4!0!} p^4 q^0 = p^4$$

$$C_4^3 p^3 q = \frac{4!}{3!1!} p^3 q = 4p^3 q$$

$$C_4^2 p^2 q^2 = \frac{4!}{2!2!} p^2 q^2 = 6p^2 q^2$$

$$C_4^1 p q^3 = \frac{4!}{1!3!} p q^3 = 4p q^3$$

$$C_4^0 p^0 q^4 = \frac{4!}{0!4!} p^0 q^4 = q^4$$

Deoarece evenimentele sunt incompatibile, *probabilitatea evenimentului reunit este egală cu suma tututror probabilităţilor fiecărui eveniment în parte, şi egală cu evenimentul sigur:*

$$C_4^4 p^4 q^0 + C_4^3 p^3 q + C_4^2 p^2 q^2 + C_4^1 p q^3 + C_4^0 p^0 q^4 =$$

$$= p^4 + 4p^3 q + 6p^2 q^2 + 4p q^3 + q^4 = 1$$

Scanned with OKEN Scanner

Aceasta înseamnă că legea de probabilitate în cazul urnei lui *Bernoulli* cu două stări este identică cu modelul matematic al dezvoltării binomului lui *Newton*. De aceea, această lege de probabilitate mai poartă şi denumirea de *legea binomială* cu repartiţia:

$$X: \begin{pmatrix} n, & n-1, & \dots, & k, & \dots, & 0 \\ p^n, & np^{n-1}q, & \dots, & C_n^k p^k q^{n-k}, & \dots, & q^n \end{pmatrix} \tag{8.3}$$

Repartiţia binomială este simetrică, dacă $p = q$. Aceasta înseamnă că termenii simetrici din cadrul dezvoltării binomului sunt egali. Dacă $p \neq q$ repartiţia binomială corespunzătoare este asimetrică. Repartiţia binomială este o repartiţie de tip discret.

Valoarea medie a variabilei în cazul repartiţiei binomiale va fi:

$$M(X) = m = \sum_{k=0}^{n} k C_n^k p^k q^{n-k} = np \tag{8.4}$$

iar dispersia

$$D^2(X) = \sigma^2 = npq \tag{8.5}$$

de unde

$$\sigma = \sqrt{npq} \tag{8.6}$$

În cazul distribuţiei binomiale, coeficientul de asimetrie β_1 şi cel de exces β_2 sunt daţi de relaţiile:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(q-p)^2}{npq} \tag{8.7}$$

şi

$$\beta_2 = \frac{\mu_4^2}{\mu_2^2} = \frac{1-6pq}{npq} + 3 \tag{8.8}$$

Aceasta înseamnă că atât β_1, cât şi β_2 sunt de ordinul $\frac{1}{n}$, iar atunci când n creşte, repartiţia binomială tinde către o *repartiţie simetrică, mezocurtică.*

Să considerăm, pentru ilustrarea aplicaţiilor practice ale repartiţiei binomiale, câteva exemple:

1. O monedă perfectă aruncată în sus de un număr mare de ori va cădea pe fiecare din cele două feţe aproximativ de acelaşi număr de ori (număr egal cu jumătate din numărul aruncărilor). La fiecare aruncare probabilitatea de a obţine *stema* este egală

Scanned with OKEN Scanner

cu probabilitatea de a obţine *valoarea* şi egală cu $\frac{1}{2}$. Pentru generalizare, să considerăm că apariţia unei feţe a monedei reprezintă un "succes", iar apariţia celeilalte feţe reprezintă un "insucces". Probabilitatea de a avea succes într-o aruncare, se notează cu p, iar probabilitatea de a obţine un insucces, cu q. Se verifică, deci, relaţiile:

$$p+q=1 \qquad \text{şi} \qquad p=q=\frac{1}{2}$$

Aceste probabilităţi rămân constante într-un număr mare de încercări independente.

2. Dacă se aruncă un zar perfect, probabilitatea de a obţine una din cele 6 feţe este egală cu $\frac{1}{6}$. În cazul aruncării zarului de un număr mare de ori, probabilitatea de a obţine faţa şaşe este egală cu probabilitatea de a obţine oricare altă faţă. Aruncările sunt independente între ele.

3. În urma observării îndelungate într-o întreprindere s-a constatat că, pentru o anumită maşină M, probabilitatea de a se deregla în decurs de o saptămână este egală cu $\frac{1}{10}$. Se cere să se determine probabilitatea ca maşina să funcţioneze normal timp de o lună şi să se reprezinte grafic repartiţia obţinută.

Tabelul 8.1. Probabilitatea $P_4(k)$

Numărul săptămânilor de funcţionare normală k	Probabilităţile $P_4(k)=C_4^k\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{4-k}$
0	$P_4(0)=C_4^0\left(\frac{9}{10}\right)^0\left(\frac{1}{10}\right)^{4-0}=0{,}0001$
1	$P_4(1)=C_4^1\left(\frac{9}{10}\right)^1\left(\frac{1}{10}\right)^3=0{,}0036$
2	$P_4(2)=C_4^2\left(\frac{9}{10}\right)^2\left(\frac{1}{10}\right)^2=0{,}0486$
3	$P_4(3)=C_4^3\left(\frac{9}{10}\right)^3\left(\frac{1}{10}\right)^1=0{,}2916$
4	$P_4(4)=C_4^4\left(\frac{9}{10}\right)^4\left(\frac{1}{10}\right)^0=0{,}6561$
Total	$\sum_{k=0}^{4}P_4(k)=\sum_{k=0}^{4}C_4^k\left(\frac{9}{10}\right)^k\left(\frac{1}{10}\right)^{4-k}=1$

Scanned with OKEN Scanner

Dacă probabilitatea de a se deregla în timp de o săptămână este egală cu $q = \frac{1}{10}$ înseamnă că probabilitatea de a funcţiona normal este egală cu $p = \frac{9}{10}$.

$$p + q = \frac{9}{10} + \frac{1}{10} = 1$$

Probabilitatea de a funcţiona normal timp de patru săptămâni consecutive se va calcula pe baza repartiţiei binomiale (tabelul 8.1).

Se observă, din tabelul 8.1, că în decursul unei luni, evenimentul cel mai probabil este funcţionarea normală a maşinii şi anume, maşina va funcţiona normal timp de patru săptămâni consecutive în 65,61% cazuri.

Altfel spus, se poate afirma cu o probabilitate P = 65,61% că maşina va funcţiona normal în decursul unei luni. Cu o probabilitate de P = 29,16% se poate afirma că maşina se va deregla în una din cele patru săptămâni; cu P = 4,86% se poate afirma că maşina se va deregla în două săptămâni consecutive dintr-o lună şi cu P = 0,36% se poate afirma că maşina se va deregla în trei saptămâni consecutive dintr-o lună şi, în sfârşit, cu o probabilitate practic egală cu zero, P = 0,01%, se poate afirma că maşina se va deregla în fiecare din cele patru săptămâni ale unei luni.

Reprezentarea grafică a repartiţiei din tabelul 8.1 se prezintă în figura 8.1.

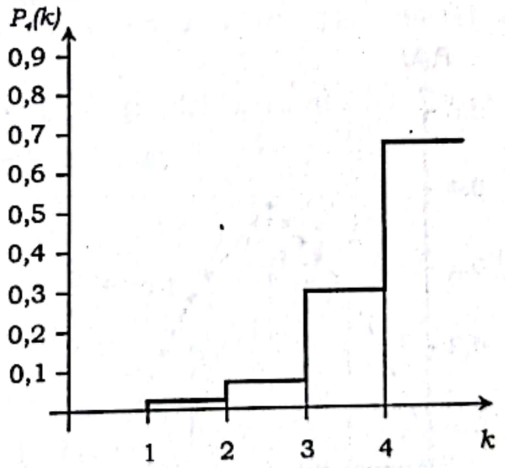


Figura 8.1. Distribuţia probabilităţii $P_4(k)$

Din figura 8.1 se observă că distribuţia este total asimetrică.

Calculul probabilităţii $P_{12}(k)$ pentru ca maşina să funcţioneze normal în decursul unui trimestru se determină în mod asemănător.

Scanned with OKEN Scanner

Să presupunem că $p = q = 0{,}15$, în exemplul anterior. În acest caz probabilităţile, ca maşina să funcţioneze normal timp de patru saptămâni consecutiv, s-au determinat în tabelul 8.2.

Tabelul 8.2. Probabilităţile $P_4(k)$

Numărul săptămânilor de funcţionare normală (k)	$P_4(k) = C_4^k 0{,}5^k 0{,}5^{4-k}$
0	$P_4(0) = C_4^0 0{,}5^0 0{,}5^4 = 0{,}0625$
1	$P_4(1) = C_4^1 0{,}5^1 0{,}5^3 = 0{,}2500$
2	$P_4(2) = C_4^2 0{,}5^2 0{,}5^2 = 0{,}3750$
3	$P_4(3) = C_4^3 0{,}5^3 0{,}5^1 = 0{,}2500$
4	$P_4(4) = C_4^4 0{,}5^4 0{,}5^0 = 0{,}0625$
Total	$\sum_{k=0}^{4} P_4(k) = \sum_{k=0}^{4} C_4^k 0{,}5^k 0{,}5^{4-k} = 1$

Distribuţia binomială obţinută în tabelul 8.2, este perfect simetrică, deoarece $p = q$. Expresia grafică a acestei distribuţii se prezintă în figura 8.2.

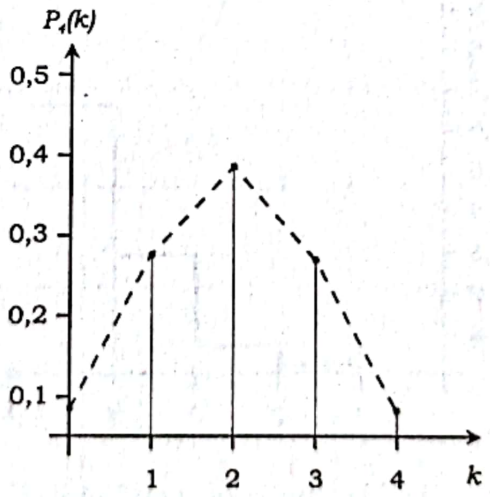


Figura 8.2. Distribuţie binomială simetrică

Scanned with OKEN Scanner

8.2. Repartiţia Poisson

Limita repartiţiei binomiale, când una dintre probabilităţi (p sau q) devine infinit de mică şi n creşte în mod corespunzător pentru a se păstra nq sau respectiv np, finit, duce la o nouă repartiţie numită *repartiţia lui Poisson*, după numele descoperitorului ei:

$$P_n(k;m) = \lim_{n\to\infty} P_n(k) = \frac{m^k}{k!}e^{-m} \qquad k = 0,1,2,\ldots,n \tag{8.9}$$

Distribuţia *Poisson* este tot o distribuţie discretă, în care fiecărei valori întregi a lui k, îi corespunde o anumită probabilitate, astfel:

$$X: \begin{pmatrix} 0 & 1 & 2 & \ldots & k & \ldots \\ e^{-m} & me^{-m} & \frac{m^2e^{-m}}{2!} & \ldots & \frac{m^ke^{-m}}{k!} & \ldots \end{pmatrix} \tag{8.10}$$

Legea de repartiţie a lui *Poisson* mai poartă şi denumirea de *legea evenimentelor rare*.

Pe baza relaţiilor (8.9) şi (8.10), rezultă că suma termenilor succesivi din dezvoltarea binomului $(p+q)^n$ este:

$$\sum_{k=0}^{m} P(k;m) = e^{-m} + me^{-m} + \frac{m^2e^{-m}}{2!} + \cdots + \frac{m^ke^{-m}}{k!} + \cdots =$$

$$= e^{-m}\left(1 + m + \frac{m^2}{2!} + \cdots + \frac{m^k}{k!} + \cdots\right) =$$

$$= e^{-m}\sum_{k=0}^{\infty}\frac{m^k}{k!} = e^{-m}\cdot e^{m} = 1$$

Scanned with OKEN Scanner

Media şi dispersia unei variabile aleatoare X, în condiţiile repartiţiei *Poisson*, sunt date de relaţiile:

$$M(X) = \sum_{k=0}^{\infty} k \frac{m^k e^{-m}}{k!} = m \tag{8.11}$$

şi

$$D^2(X) = M[X - M(X)]^2 = \sigma^2 \tag{8.12}$$

Relaţiile (8.11) şi (8.12), arată că media şi dispersia unei variabile aleatoare ce urmează repartiţia *Poisson* sunt egale între ele şi în acelaşi timp egale cu parametrul repartiţiei.

Această proprietate prezintă o importanţă deosebită în practică pentru alegerea tipului de repartiţie teoretică, care corespunde cel mai bine unei distribuţii empirice date. Astfel, dacă în cazul unei distribuţii empirice media şi dispersia variabilei vor fi egale sau apropiate ca mărime, aceasta va constitui un argument în favoarea ipotezei că repartiţia empirică se apropie de o repartiţie *Poisson* şi invers.

Indicatorii de asimetrie şi exces, în cazul repartiţiei *Poisson* sunt daţi de expresiile:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{m^2}{m^3} = \frac{1}{m} \tag{8.13}$$

şi

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3m^2 + m}{m^2} = 3 + \frac{1}{m} \tag{8.14}$$

Relaţiile (8.13) şi (8.14) arată că β_1 şi β_2 reprezintă limitele expresiei $(p+q)^n$, când $P = \frac{m}{n}$ şi n tinde către infinit.

Pentru ilustrarea modului de aplicare în practică a acestei distribuţii, să considerăm un exemplu. O maşină automată efectuează o anumită operaţie la piese de acelaşi tip, în serii mari. Pentru controlul modului de funcţionare a maşinii se extrage un eşantion de volum n = 100 piese. Probabilitatea de a extrage o piesă rebut este de 0,01. Se cere să se determine probabilitatea ca în eşantionul de 100 piese să se găsească o piesă rebut, două piese rebut şi nici o piesă rebut.

Media variabilei empirice este: $m = np = 100 \cdot 0{,}01 = 1$.

Din tabelele cu repartiţia *Poisson* rezultă probabilităţile corespunzătoare:

- pentru o piesă rebut: $P_1 = 0{,}368$

Scanned with OKEN Scanner

- pentru două piese rebut: $P_2 = 0{,}184$
- pentru zero piese rebut: $P_0 = 0{,}368$

La aceleaşi rezultate se poate ajunge şi aplicând formula (8.9).

8.3. Repartiţia normală

Descoperită de către matematicianul englez de origine franceză, *Abraham de Moivre*, *legea normală* ocupă un loc central în cadrul teoriei şi practicii statistice şi reprezintă forma continuă, limită a unei distribuţii binomiale, când n tinde către infinit. Contribuţii deosebit de importante în caracterizarea distribuţiei normale au avut şi alţi oameni de ştiinţă, între care un loc mai important deţin matematicianul, astronomul şi fizicianul francez *Pierre-Simon Laplace* (1749-1827) şi matematicianul, fizicianul şi astronomul german *Karl-Friedrich Gauss* (1775-1855). De aceea, expresia geometrică a distribuţiei mai poartă şi denumirea de curba lui Gauss sau a lui Gauss-Laplace, denumire ce nu este întru totul îndreptăţită.

Distribuţia normală se notează $n(x, m, \sigma)$, pentru a pune în evidenţă că, în afară de argumentul x, depinde şi de parametrii m şi σ:

$$n(x, m, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{8.15}$$

în care $x \in (-\infty; +\infty)$

Parametrul m reprezintă media variabilei x, iar σ, abaterea medie pătratică. Aceasta înseamnă că funcţia de repartiţie a unei variabile normale este perfect determinată de m şi σ.

Curba (8.15) reprezintă forma de bază a distribuţiei normale şi delimitează faţă de axa ox o suprafaţă egală cu unitatea. În formula (8.15), totalul frecvenţelor variabilei este considerat egal cu 1 şi de aceea, pentru determinarea ordonatelor în cazuri particulare, concrete, frecvenţele relative obţinute prin aplicarea formulei (8.15) se înmulţesc cu volumul colectivităţii N.

Ecuaţia generală a curbei normale are următoarea expresie:

$$n(x, m, \sigma) = \frac{N}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{8.16}$$

Scanned with OKEN Scanner

Funcţia $n(x,m,\sigma)$ este o *densitate de probabilitate*, deoarece îndeplineşte următoarele condiţii:

1. $n(x,m,\sigma)\geq 0$. Această condiţie este evidentă, deoarece funcţia $n(x,m,\sigma)$ este formată numai din elemente pozitive, iar când x tinde către $\pm\infty$, $n(x,m,\sigma)$ tinde către zero.

$$\int_{-\infty}^{+\infty} n(x,m,\sigma)dx = 1 \tag{8.17}$$

sau

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{-\frac{(x-m)^2}{2\sigma^2}}dx = 1 \tag{8.18}$$

Media variabilei aleatoare X, care admite pe $f(x)$ ca densitate de probabilitate, este dată de expresia:

$$M(X)=\int_{-\infty}^{+\infty} xn(x,m,\sigma)dx=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{+\infty} x\,e^{-\frac{(x-m)^2}{2\sigma^2}}dx = m \tag{8.19}$$

rezultă

$$M(X)=m \tag{8.20}$$

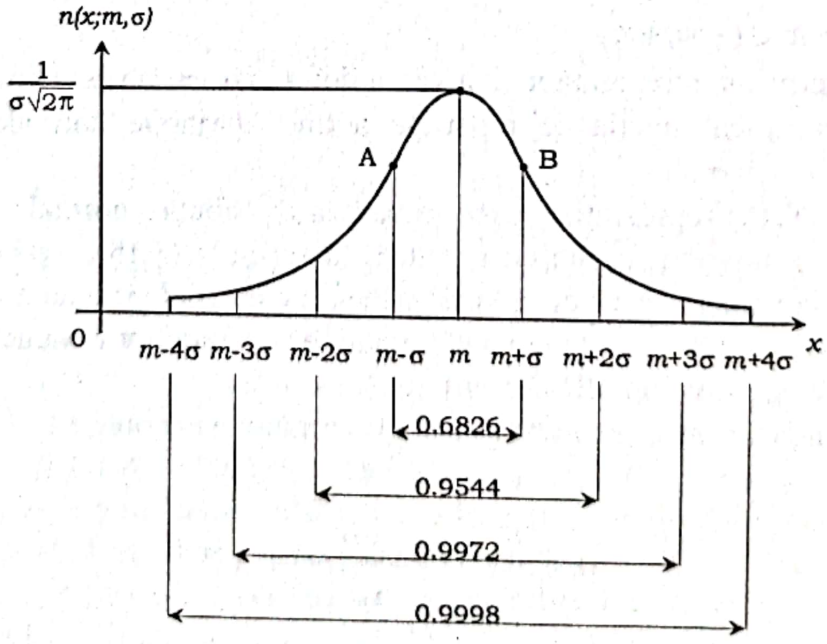


Figura 8.3. Curba normală

Scanned with OKEN Scanner

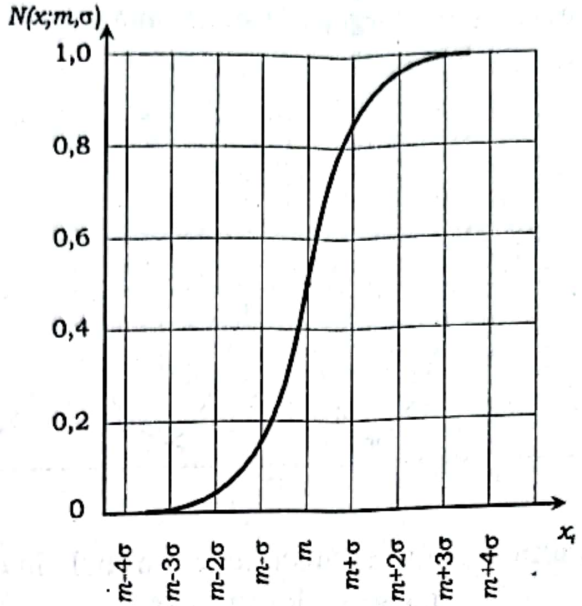


Figura 8.4. Curba normală cumulată

Dispersia variabilei aleatoare X va fi:

$$D^2(X) = \int_{-\infty}^{+\infty} [x - M(X)]^2 f(x)dx = \sigma^2$$

de unde

$$D(X) = \sigma \tag{8.21}$$

8.3.1. Proprietăţile distribuţiei normale

1. *Distribuţia normală este unimodală, simetrică faţă de dreapta* $x = m$ *şi continuă.* Din punct de vedere teoretic, amplitudinea ei acoperă intervalul $(-\infty,+\infty)$. În realitate, 68,3% din aria mărginită de curba normală şi *ox*, se găseşte între ordonatele ce trec prin punctele de abscisă $x-\sigma$ şi $x+\sigma$; 95,4% se află între $x-2\sigma$ şi $x+2\sigma$; 99,72% se situează între $x-3\sigma$ şi $x+3\sigma$ etc.

2. *Punctele, de pe curbă, a căror abscisă este egală cu* $x \pm \sigma$ *sunt puncte de inflexiune*, respectiv punctele A şi B din figura 8.3.

Scanned with OKEN Scanner

Forma curbei normale depinde de cei doi parametri m şi σ, iar modificarea ei, sub influenţa variaţiei celor doi parametri, se prezintă sugestiv în figura 8.5.

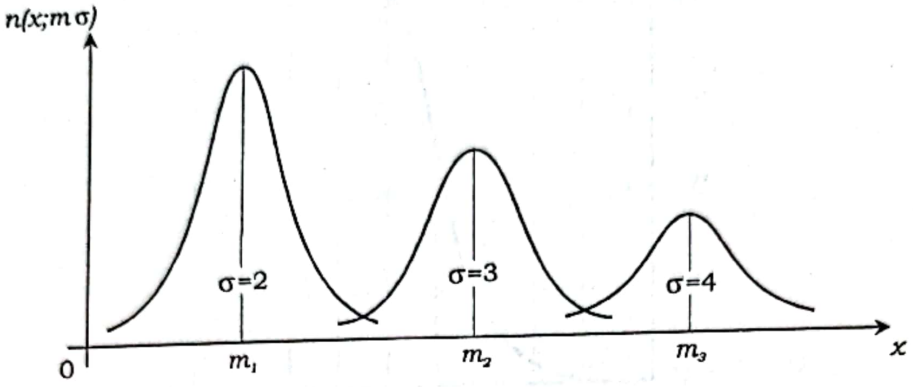


Figura 8.5. Modificarea formei distribuţiei normale în cazul variaţiei parametrilor m şi σ

Creşterea gradului de împrăştiere a variantelor variabilei X în jurul nivelului lor mediu, înseamnă reducerea ordonatei maxime, concomitent cu mărirea amplitudinii variaţiei şi invers, în timp ce creşterea sau reducerea nivelului mediu al variaţiei de la o distribuţie la alta, determină doar o mişcare de translaţie a curbei pe axa *ox*. Mărimea ordonatei valorii modale a variabilei este egală cu $\frac{1}{\sigma\sqrt{2\pi}}$, ceea ce înseamnă că ea depinde numai de valoarea abaterii standard.

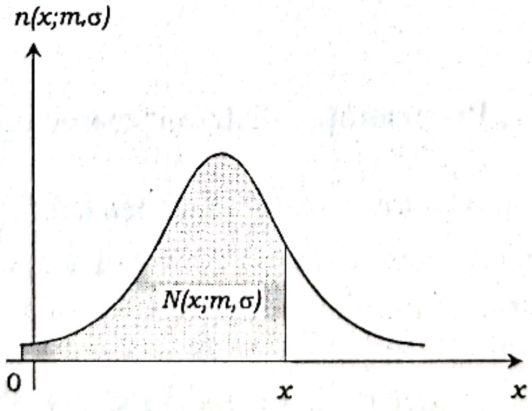


Figura 8.6. Curba de repartiţie normală

Scanned with OKEN Scanner

3. *În orice distribuţie normală, media aritmetică, mediana şi modul coincid* datorită faptului că distribuţia este perfect simetrică.

4. *Momentele centrate de ordin impar sunt nule*, întrucât abaterile $(x_i - \bar{x})$, care intră în calculul lor, îşi păstrează semnul. Este de menţionat faptul că *momentele centrate de ordinul întâi sunt nule pentru toate tipurile de distribuţii* (simetrice sau asimetrice).

Distribuţia normală centrată se obţine făcând o mişcare de translaţie a curbei pe axa x, astfel încât punctul de abscisă m să coincidă cu originea sistemului de axe rectangulare, ceea ce înseamnă că $m = 0$.

Făcând substituţia $m = 0$ în (8.15), se obţine *distribuţia normală centrată*:

$$n(x,0,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2\sigma^2}} \tag{8.22}$$

Dacă în formula densităţii de probabilitate (8.16) se face schimbarea de variabilă $\frac{x-m}{\sigma\sqrt{2}} = t$ se obţine *funcţia de repartiţie* $(F(x))$, notată cu $N(x,m,\sigma)$:

$$F(x) = N(x,m,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{(t-m)^2}{2\sigma^2}}\, dt \tag{8.23}$$

unde $x = m + t\sigma\sqrt{2}$ iar $dx = \sigma\sqrt{2}\, dt$

Funcţia de repartiţie exprimă probabilitatea ca o variabilă aleatoare X să ia o valoare mai mică decât o valoare dată x, astfel:

$$F(x) = P(X < x) = N(x,m,\sigma)$$

Expresia geometrică a funcţiei (8.23) se prezintă în figura 8.6.

Făcând transformarea $\frac{t-m}{\sigma} = y$ şi schimbând limitele de integrare în formula (8.23), rezultă:

$$P(X < x) = N(x,m,\sigma) \text{ şi } P(X < m + z\sigma) = N(m + \sigma z; m, \sigma)$$

Scanned with OKEN Scanner

$$F(X<x)=N(z,0,1)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{z}e^{-\frac{y^2}{2}}dy=N(z,0,1) \tag{8.24}$$

unde $t=m+\sigma y$ şi $dt=\sigma\,dy$

Funcţia (8.24) poartă denumirea de *funcţie de repartiţie normală normată.* Probabilitatea ca X să fie cuprins în intervalul $(a\,,\,b)$ este dată de relaţia:

$$P(a<X<b)=\frac{1}{\sigma\sqrt{2\pi}}\int_{\frac{a-m}{\sigma}}^{\frac{b-m}{\sigma}}e^{-\frac{y^2}{2}}dy \tag{8.25}$$

sau notând $\frac{a-m}{\sigma}=z_1$ şi $\frac{b-m}{\sigma}=z_2$

rezultă: $$P(a<X<b)=\frac{1}{\sigma\sqrt{2\pi}}\int_{z_1}^{z_2}e^{-\frac{y^2}{2}}dy \tag{8.26}$$

Funcţia integrală a lui *Laplace*, simetrică faţă de origine, este dată de relaţia:

$$\Phi(z)=\frac{1}{\sigma\sqrt{2\pi}}\int_{0}^{z}e^{-\frac{y^2}{2}}dy \tag{8.27}$$

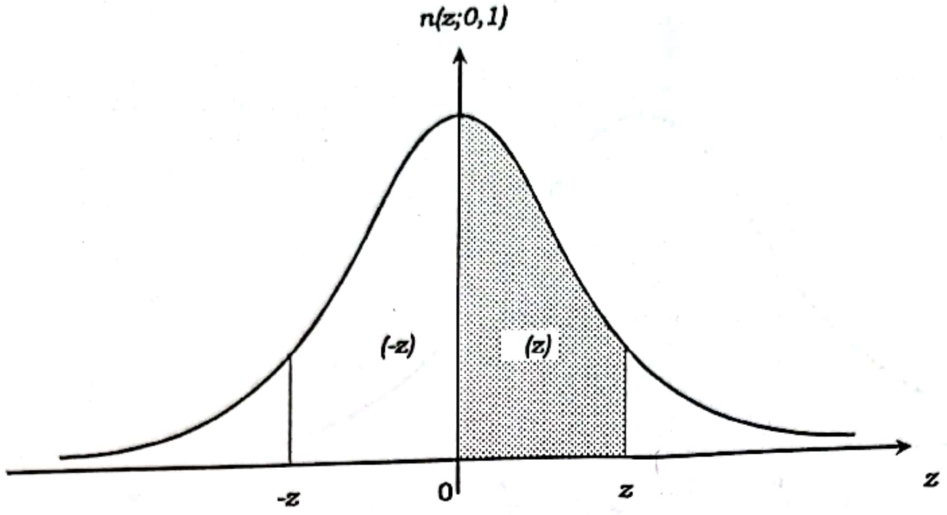
şi $$\Phi(-z)=-\Phi(z) \tag{8.28}$$

Această funcţie (fig.8.7) prezintă o importanţă deosebită în practica statistică şi de aceea i s-au calculat valorile în tabele speciale.

Cu ajutorul funcţiei lui *Laplace* $\Phi(z)$, funcţia de repartiţie normată (8.24), devine:

$$N(z,0,1)=\frac{1}{2}+\Phi(z) \tag{8.29}$$

Scanned with OKEN Scanner



Valori mai importante ale funcţiei lui *Laplace*

$\Phi(0) = 0$

$\Phi(-\infty) = -\frac{1}{2}$

$\Phi(+\infty) = \frac{1}{2}$

Figura 8.7. Curba lui Laplace

iar funcţia de repartiţie nenormată va fi:

$$N(x, m, \sigma) = \frac{1}{2} + \Phi\left(\frac{x-m}{\sigma}\right) \tag{8.30}$$

Pe baza relaţiilor (8.29) şi (8.30), probabilităţile (8.23) şi (8.26) se exprimă astfel:

$$P(X < x) = \frac{1}{2} + \Phi\left(\frac{x-m}{\sigma}\right) \tag{8.31}$$

şi

$$P(a < X < b) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right) \tag{8.32}$$

Probabilitatea (8.32) se ilustrează grafic în figura 8.8.

Scanned with OKEN Scanner

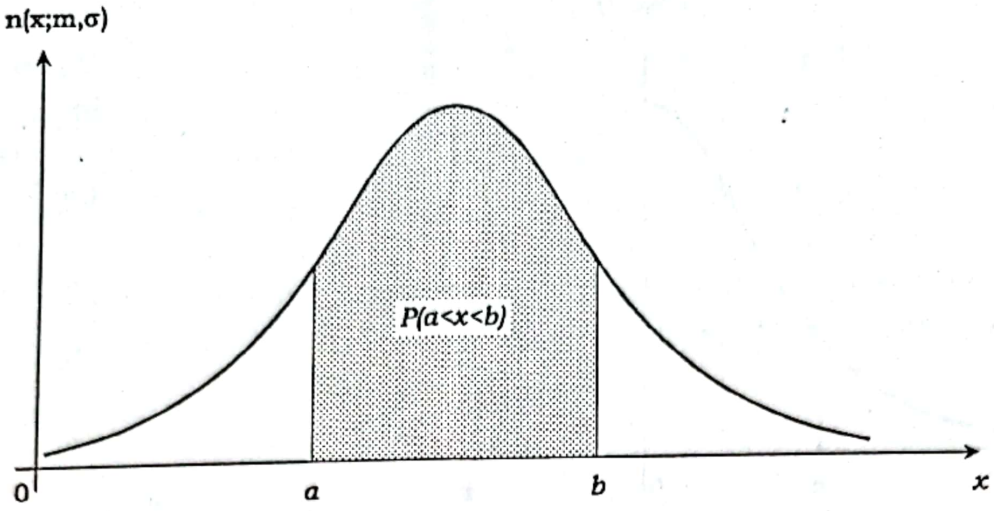


Figura 8.8. Probabilitatea ca $X \in (a,b)$

Exemplu. Fineţea firelor de bumbac produse de o filatură urmează o distribuţie normală cu parametrii $m = 50$ şi $\sigma = 1{,}5$.

a) Să se calculeze probabilitatea ca fineţea să ia valori mai mici decât 46.

b) Să se calculeze probabilitatea ca fineţea firelor lotului să se cuprindă în intervalele (46;54) şi (46;53).

Soluţie

a) Probabilitatea ca fineţea firelor din lot să ia o valoare mai mică decât 46 este dată de relaţia (8.31):

$$P(X < 46) = \frac{1}{2} + \Phi\left(\frac{46-50}{1{,}5}\right) = \frac{1}{2} + \Phi(-2{,}66)$$

Din tabelele cu funcţia integrală a lui *Laplace* $\Phi(z)$ rezultă pentru $z = 2{,}66$ că $\Phi(2{,}66) = 0{,}49609$, iar $\Phi(-z) = -\Phi(z)$, deci $\Phi(-2{,}66) = -0{,}49609$.

Rezultă că $P(X < 46) = \frac{1}{2} - 0{,}49609 = 0{,}00391$, ceea ce înseamnă că numai în 0,4% din cantitatea de fire cuprinse în lot fineţea va fi inferioară lui 46.

b) Probabilitatea ca fineţea firelor de bumbac din lot să se cuprindă într-un interval este dată de relaţia (8.23):

Scanned with OKEN Scanner

$$P(46 < X < 54) = \Phi\left(\frac{54-50}{1,5}\right) - \Phi\left(\frac{46-50}{1,5}\right) = \Phi(2,66) - \Phi(-2,66) =$$

$$= 2 \cdot 0,49609 = 0,992$$

iar

$$P(47 < X < 53) = \Phi\left(\frac{53-50}{1,5}\right) - \Phi\left(\frac{47-50}{1,5}\right) = 2\Phi(2) =$$

$$= 2 \cdot 0,47725 = 0,95450$$

Deci, se poate afirma că finețea firelor din lot se va cuprinde în intervalul (46;54) într-o proporţie de 99,3%, iar dacă se micşorează intervalul de încredere, se micşorează şi probabilitatea cu care se garantează rezultatele obţinute.

Repartiţia normală ocupă un loc important în statistică deoarece:

- majoritatea măsurătorilor fenomenelor naturale şi a celor sociale sunt repartizate aproximativ normal;
- variabilele care nu sunt normal repartizate prin anumite transformări duc la noi variabile repartizate exact sau aproximativ normal;
- distribuţia sumei unor variabile aleatoare este aproximativ normal repartizată indiferent de forma distribuţiilor fiecărei variabile în parte.

Scanned with OKEN Scanner

9. CARACTERISTICI ALE FORMEI DISTRIBUŢIILOR

9.1. Asimetria distribuţiilor

Distribuţiile empirice pot să difere între ele în funcţie de următoarele elemente: valoarea medie, dispersia, gradul de asimetrie şi exces sau combinaţia dintre ele. O distribuţie a unei variabile X cu variantele x_i şi frecvenţele $f(x_i)$ este numită simetrică faţă de o valoare medie m, dacă este satisfăcută relaţia:

$$f(m-\Delta)=f(m+\Delta) \qquad \text{pentru orice} \qquad \Delta = x - m\,.$$

Diferenţele dintre distribuţii din punct de vedere al formei pot fi mai bine puse în evidenţă prin reprezentarea lor grafică. Astfel, în figura 9.1, se prezintă două *distribuţii simetrice*, care diferă între ele numai prin valorile lor medii, având aceeaşi dispersie.

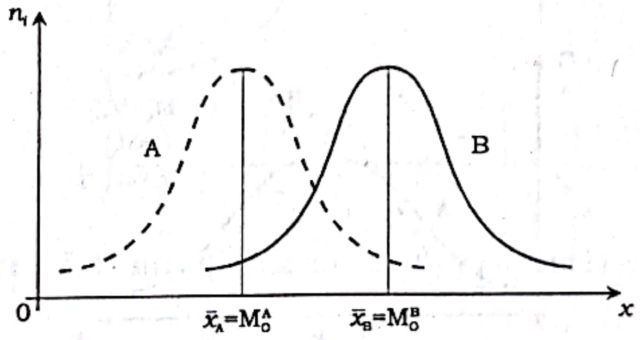


Figura 9.1. Distribuţii normale cu medii diferite şi aceeaşi dispersie

Distribuţia B este localizată mai la dreapta distribuţiei A şi ambele sunt simetrice faţă de perpendicularele ridicate din punctele de abscisă $\bar{x}_A$ şi $\bar{x}_B$. Distribuţia B are media mai mare decât distribuţia A:

$$\bar{x}_A < \bar{x}_B$$

Scanned with OKEN Scanner

Alte tipuri de curbe simetrice se prezintă în figura 9.2.
Aceste distribuţii au ca axe de simetrie dreptele d şi respectiv d'.

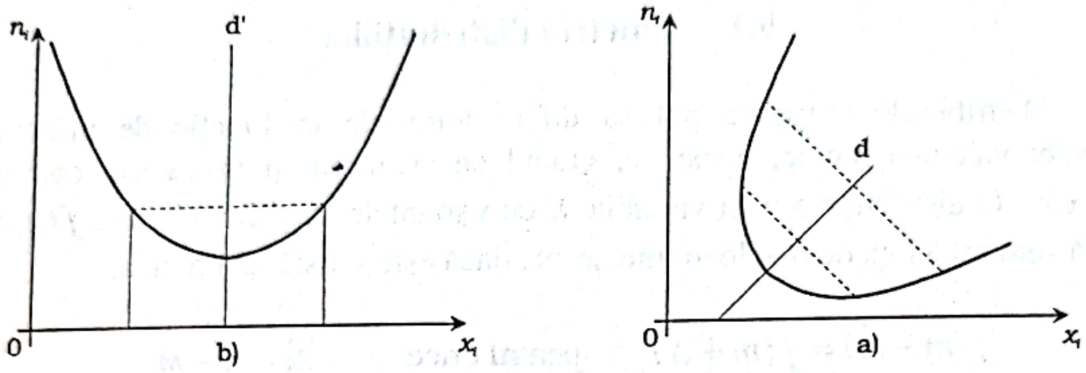


Figura 9.2. Distribuţii simetrice

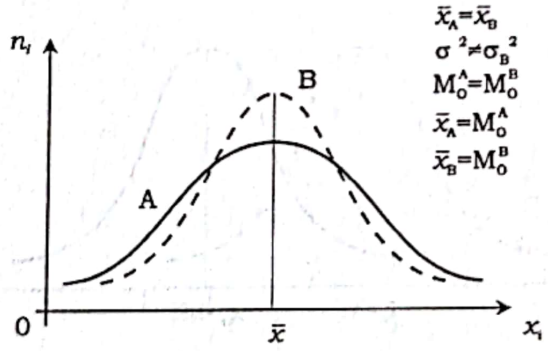


Figura 9.3. Distribuţii normale simetrice cu grade diferite de dispersie

În figura 9.3 se prezintă două distribuţii simetrice, care au aceeaşi medie, dar diferă prin dispersie. Distribuţia A are o dispersie mai mare decât distribuţia B şi, în consecinţă, o *omogenitate* mai redusă a variantelor variabilei X. Distribuţia B prezintă un exces de valori ale varibilei vecine cu media, este deci mai boltită decât *distribuţia normală*.

Scanned with OKEN Scanner

În figura 9.4 se prezintă o *distribuţie asimetrică*, care are mai multe valori extrem de mari şi mai puţine extrem de mici comparativ cu distribuţa normală. Distribuţia *B* poartă denumirea - convenţional stabilită - de *distribuţie unimodală cu asimetrie la dreapta* sau distribuţie cu *asimetrie pozitivă*.

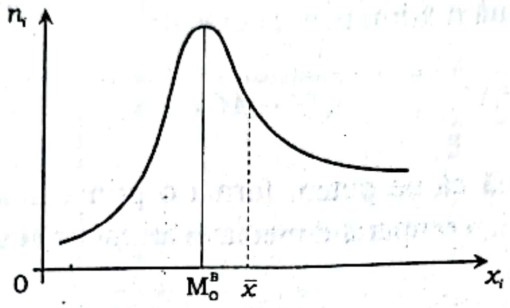


Figura 9.4. Distribuţie unimodală cu asimetrie la dreapta

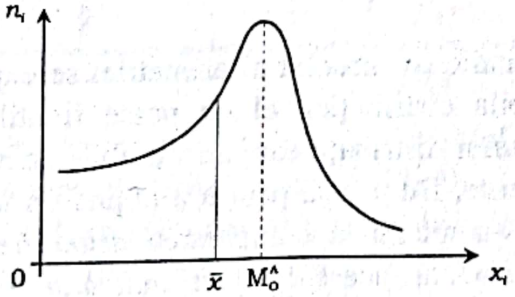


Figura 9.5. Distribuţia unimodală cu asimetrie la stânga

Figura 9.5 prezintă un caz de *distribuţie unimodală cu asimetrie la stânga*, sau de distribuţie cu *asimetrie negativă*. În această distribuţie se găsesc mai multe variante extrem de mici ale variabilei şi mai puţine variante extrem de mari, decât în cazul distribuţiei normale cu aceeaşi medie şi dispersie.

Din punct de vedere al formei lor, distribuţiile statistice pot fi *simetrice* sau *asimetrice*. La rândul lor, distribuţiile asimetrice se deosebesc între ele prin gradul şi sensul de asimetrie pe care-l prezintă.

În literatura de specialitate s-au propus mai mulţi indicatori pentru măsurarea gradului de *asimetrie* (sau de *oblicitate*) a unei distribuţii empirice, dar nici unul nu

Scanned with OKEN Scanner

este unanim acceptat. De aceea, ori de câte ori se prezintă un indicator de asimetrie, este necesar să se precizeze şi metoda care a stat la baza calculării lui.

Există două categorii de *indicatori ai asimetriei* calculaţi prin compararea, sub formă de diferenţă sau de raport, a doi dintre indicatorii tendinţei centrale şi alţii bazaţi pe raportul dintre momente de diferite ordine.

Asimetria unei distribuţii va fi cu atât mai pronunţată, cu cât diferenţa dintre *media aritmetică* şi *mod* va fi mai mare şi invers. În cazul distribuţiilor unimodale perfect simetrice, cele două mărimi medii coincid:

$$\bar{x} = Mo$$

Aceasta înseamnă că ne putem forma o primă imagine asupra *gradului de asimetrie* a unei distribuţii, comparând media ei aritmetică cu modul, pe baza relaţiei:

$$\Delta = \bar{x} - Mo \tag{9.1}$$

unde
$$\begin{cases} \Delta > 0 & \textit{pentru asimetrie pozitivă} \\ \Delta < 0 & \textit{pentru asimetrie negativă} \\ \Delta = 0 & \textit{pentru simetrie} \end{cases}$$

Deoarece acest indicator absolut al asimetriei se exprimă în aceleaşi unităţi de măsură ca şi variabila distribuţiei, el nu poate fi utilizat pentru compararea gradului de asimetrie al unor distribuţii care au variabile de natură diferită, exprimate în unităţi de măsură diferite. De aceea, pentru a obţine un indicator al asimetriei cu aplicabilitate generală, s-a recurs la compararea *asimetriei absolute* cu *abaterea standard* σ, obţinându-se în acest fel, un *indicator relativ (standardizat) al asimetriei*, propus de către *Karl Pearson* (1857-1936), statistician, biolog şi filozof englez:

$$S_k = \frac{\bar{x} - Mo}{\sigma} \tag{9.2}$$

Din această formulă rezultă că, dacă:

$S_k > 0$, *distribuţia prezintă asimetrie pozitivă;*
$S_k < 0$, *distribuţia prezintă asimetrie negativă;*
$S_k = 0$, *distribuţia este simetrică.*

Scanned with OKEN Scanner

Indicatorii relativi de asimetrie sunt mărimi relative abstracte (independente de unitatea de măsură a caracteristicii) şi exprimă măsura în care o distribuţie oarecare se abate de la simetrie şi direcţia asimetriei.

Să considerăm, pentru exemplificarea acestui indicator, distribuţia mamelor din *România*, după vârstă în anul 1990 (tatăl făcând parte din grupa de vârstă 50 ani şi peste, tabelul 9.1).

Tabelul 9.1. Calculul simplificat al mediei şi dispersiei

Vârsta mamei (ani)	Numărul mamelor n_i	x_i	$x_i - 37{,}5$	$\frac{x_i - 37{,}5}{5}$	$x_i' n_i$	$x_i'^2 n_i$	N_i
15-19	16	17,5	-20	-4	-64	256	16
20-24	38	22,5	-15	-3	-114	342	54
25-29	86	27,5	-10	-2	-172	344	140
30-34	182	32,5	- 5	-1	-182	182	322
35-39	421	37,5	0	0	0	0	743
40-44	518	42,5	5	1	518	518	1261
45-49	67	47,5	10	2	134	268	1328
Total	1328	-	-	-	120	1910	-

Sursa: Prelucrat după *Anuarul Statistic al României,1991,CNS, Bucureşti, p.70-71*

Asimetria distribuţiei din tabelul 9.1, ca de altfel a oricărei distribuţii, va deveni mai evidentă prin reprezentarea ei grafică (fig. 9.6). Din grafic se observă că există un număr mai mare de mame care au o vârstă mai apropiată de cea minimă a fertilităţii şi un număr mai mic de mame care se apropie de vârsta maximă fertilă, decât în cazul distribuirii lor după o lege normală de repartiţie. Deci, distribuţia mamelor pe grupe de vârstă prezintă o asimetrie negativă. Gradul de asimetrie va fi determinat cu ajutorul *coeficientului de asimetrie* al lui *Pearson* (9.1).

$$\bar{x} = \frac{\sum_{i=1}^{k} \frac{x_i - a}{b} n_i}{\sum_{i=1}^{k} n_i} b + a = \frac{120}{1328} \cdot 5 + 37{,}5 = 37{,}95 \; ani$$

$$Mo = x_{i-1} + d \frac{\Delta_1}{\Delta_1 + \Delta_2} = 40 + 5 \cdot \frac{97}{97 + 451} = 40{,}89 \; ani$$

Scanned with OKEN Scanner

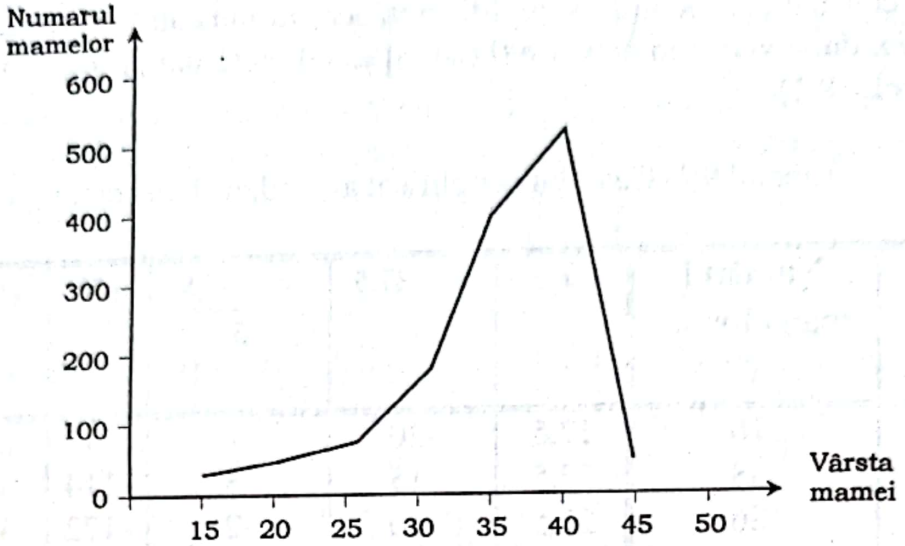


Figura 9.6. Distribuţia mamelor după vârstă

$$\sigma = \sqrt{\frac{\sum_{i=1}^{k}\left(\frac{x_i' - a}{b}\right)^2 n_i}{\sum_{i=1}^{k} n_i} b^2 - (\bar{x} - a)^2} = \sqrt{\frac{1910}{1328} \cdot 25 - (37{,}95 - 37{,}5)^2} = 6{,}31 \; ani$$

Aplicând formula (9.2), rezultă:

$$S_k = \frac{\bar{x} - Mo}{\sigma} = \frac{37{,}95 - 40{,}89}{6{,}31} = -0{,}46$$

$S_k = -0{,}46$ confirmă faptul că distribuţia mamelor după vârstă prezintă o asimetrie negativă moderată.

Deoarece în cazul seriilor de intervale valoarea modului nu poate fi determinată în mod precis, diferite metode de grupare dând de regulă, rezultate diferite, formula coeficientului de asimetrie *Pearson* (9.2) nu este întru totul satisfăcătoare.

În cazul unor distribuţii cu frecvenţă, moderat asimetrice, între *mod, mediană* şi *media aritmetică* există următoarea *relaţie aproximativă*:

Scanned with OKEN Scanner

$$Mo = 3Me - 2\bar{x} \tag{9.3}$$

Substituind această valoare a modului în formula (9.2), rezultă:

$$S_k = \frac{\bar{x} - 3Me + 2\bar{x}}{\sigma} = \frac{3(\bar{x} - Me)}{\sigma} \tag{9.4}$$

iar $$-3 \le S_k \le 3$$

Desigur, această nouă expresie a lui S_k înlătură parţial deficienţa coeficientului de asimetrie (9.2), legată de determinarea aproximativă a modului.

Calculul şi analiza coeficientului de asimetrie cu ajutorul formulei (9.4) se poate exemplifica pe baza distribuţiei din tabelul 9.1:

$$Me = x_{i-1} + d\frac{U^{me} - N_{i-1}}{n_i} = 35 + 5 \cdot \frac{664 - 322}{421} = 39{,}06 \; ani$$

$$S_k = \frac{3(\bar{x} - Me)}{\sigma} = \frac{3(37{,}95 - 39{,}06)}{6{,}31} = -0{,}52$$

Coeficientul de asimetrie $S_k = -0{,}52$ arată că distribuţia mamelor pe grupe de vârstă prezintă o asimetrie negativă moderată.

Într-o distribuţie simetrică, cuartelele Q_1 şi Q_2 se situează de o parte şi de alta a medianei, în mod simetric, adică se verifică relaţia: $Me - Q_1 = Q_3 - Me$. Dacă distribuţia este asimetrică, cuartilele nu vor mai fi echidistante. Pornind de la acest fapt, *Arthur Bowley* (1869-1957), statistician, matematician şi economist englez, a propus pentru calculul coeficientului de asimetrie, următoarea formulă:

$$S_k = \frac{(Q_3 - Me) - (Me - Q_1)}{Q_3 - Q_1} \tag{9.5}$$

sau $$S_k = \frac{Q_1 + Q_3 - 2Me}{Q_3 - Q_1} \tag{9.6}$$

Scanned with OKEN Scanner

Din formula (9.6) se observă că:

$S_k = 0$, când $Me - Q_1 = Q_3 - Me$ adică cuartilele sunt echidistante faţă de mediană, ceea ce se întâmplă în cazul *distribuţiilor simetrice*;

$S_k > 0$, dacă distribuţia prezintă o *asimetrie pozitivă*, adică $(Q_3 - Me) > (Me - Q_1)$;

$S_k < 0$, dacă $(Me - Q_1) > (Q_3 - Me)$, în cazul distribuţiilor cu *asimetrie negativă*.

Împărţind numărătorul şi numitorul relaţiei (9.6) prin 2 rezultă:

$$S_k = \frac{\frac{Q_1 + Q_3}{2} - Me}{\frac{Q_3 - Q_1}{2}}$$

notând $k = \frac{Q_1 + Q_3}{2}$ şi $k' = \frac{Q_3 - Q_1}{2}$

se obţine $$S_k = \frac{k - Me}{k'} \qquad (9.7)$$

Aceasta înseamnă că, în general, coeficientul de asimetrie a lui *Bowley* are acelaşi semn şi aceeaşi semnificaţie ca şi cel al lui *Pearson*. Deoarece mediana se situează întotdeauna între cuartilele Q_1 şi Q_3, se pot distinge două cazuri extreme:

$$S_k = \frac{Q_3 + Q_1 - 2Q_1}{Q_3 - Q_1} = \frac{Q_3 - Q_1}{Q_3 - Q_1} = 1$$

$$S_k = \frac{Q_3 + Q_1 - 2Q_3}{Q_3 - Q_1} = -\frac{Q_3 - Q_1}{Q_3 - Q_1} = -1$$

Coeficientul de asimetrie al lui *Bowley* variază între -1 şi +1. În cazul asimetriei pozitive maxime, $Me = Q_1$ iar $S_k = 1$, iar pentru o distribuţie cu o asimetrie negativă maximă $S_k = -1$ şi $Me = Q_3$. Deficienţa principală a acestui

Scanned with OKEN Scanner

coeficient este dată de faptul că neglijează prima şi ultima pătrime a variantelor unei caracteristici dintr-o serie ordonată.

Coeficientul de asimetrie al lui *Bowley* poate fi calculat şi pe baza a două decile sau a două centile echidistante faţă de mediană, de exemplu D_1 şi D_9; C_{10} şi C_{90} sau C_1 şi C_{99} etc.

În cazul utilizării abaterilor centilice echidistante faţă de mediană, indicatorul asimetriei va fi:

$$S_k = \frac{(C_{90} - Me) - (Me - C_{10})}{C_{90} - C_{10}} \quad \text{sau} \quad S_k = \frac{C_{10} + C_{90} - 2Me}{C_{90} - C_{10}} \tag{9.8}$$

T. Kelley a propus un coeficient de asimetrie, mai mult cu o semnificaţie teoretică, calculat pe baza centilelor C_{10} şi C_{90}, după următoarea formulă:

$$S_k = C_{50} - \frac{C_{90} + C_{10}}{2} \tag{9.9}$$

Un alt indicator al asimetriei se poate obţine pe baza momentelor centrate de ordinul doi şi trei.

Astfel, în cazul unei distribuţii simetrice, după cum se ştie, *momentele centrate de ordin impar sunt nule. Momentul centrat de ordinul 1 este nul pentru orice distribuţie (simetrică sau asimetrică).* Pornind de la aceste considerente, *K. Pearson* a propus şi următorul coeficient de asimetrie:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} \tag{9.10}$$

Din formula (9.10) rezultă că β_1 este un indicator standardizat al asimetriei, fără dimensiune, invariabil prin schimbarea originii şi nul pentru distribuţii simetrice. În cazul unei distribuţii simetrice, $\beta_1 = 0$, iar în cazul distribuţiilor asimetrice, $\beta_1 > 0$. Sensul asimetriei este dat de semnul lui μ_3 şi anume, $\mu_3 > 0$ în cazul distribuţiilor cu *asimetrie la dreapta* şi $\mu_3 < 0$, pentru distribuţiile cu *asimetrie negativă.*

Calculul preliminar al momentelor centrate de ordinul doi şi trei pentru distribuţia din tabelul 9.1 se prezintă în tabelul 9.2.

Scanned with OKEN Scanner

$$\mu_2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{47516,89}{1328} = 35,78$$

$$\mu_3 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^3 n_i}{\sum_{i=1}^{k} n_i} = \frac{-107149,64}{1328} = -80,68$$

Tabelul 9.2. Calculul momentelor centrate de ordinul doi, trei şi patru

Vârsta mamei	n_i	x_i	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_i$	$(x_i - \bar{x})^3 n_i$	$(x_i - \bar{x})^4 n_i$
15-19	16	17,5	418,20	6691,24	-136835,86	2798276,57
20-24	38	22,5	238,70	9070,70	-140142,24	2165197,57
25-29	86	27,5	109,20	9391,42	-98140,29	1025566,00
30-34	182	32,5	29,70	5405,86	-29461,91	160567,41
35-39	421	37,5	0,20	85,25	-38,36	17,26
40-44	518	42,5	20,70	10723,90	48793,72	222011,44
45-49	67	47,5	91,20	6110,57	58355,92	557299,03
Total	1328	-	-	47516,89	-107149,64	6928935,28

Substituind aceste date în formula (9.10), rezultă:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{(-80,68)^2}{35,78^3} = \frac{6510,06}{1315,99} = 4,94$$

$\beta_1 = 4,94$ evidenţiază faptul că distribuţia mamelor după vârstă prezintă o asimetrie negativă moderată, deoarece $\mu_3 = -80,68 < 0$.

Scanned with OKEN Scanner

9.2 Boltirea distribuţiilor

Indicatorii boltirii sau *excesului* arată dacă o distribuţie este mai boltită sau mai puţin boltită decât distribuţia normală. Dacă într-o distribuţie variantele unei variabile sunt concentrate în jurul nivelului ei mediu mai mult decât în cazul unei distribuţii normale, atunci distribuţia considerată va fi mai boltită decât aceasta, iar graficul ei poartă denumirea de *curbă leptocurtică*. Dacă curba este mai puţin boltită decât cea normală, ea poartă denumirea de *curbă platicurtică*. Curba normală este numită *curbă mezocurtică*.

Pentru măsurarea gradului de boltire a distribuţiilor, se utilizează *coeficientul de boltire K. Pearson*, dat de formula:

$$\beta_2 = \frac{\mu_4}{\mu_2^2} \tag{9.11}$$

în care $\mu_4 = \dfrac{\sum_{i=1}^{k}(x_i - \bar{x})^4 n_i}{\sum_{i=1}^{k} n_i}$ iar $\mu_2 = \sigma^2$

Se ştie că, momentul centrat de ordinul s, notat cu μ_s în cazul unei distribuţii normale, satisface următoarea relaţie de recurenţă, care permite exprimarea momentelor de ordin superior în funcţie de momentele de ordin inferior:

$$\mu_s = (s-1)\sigma^2 \cdot \mu_{s-z} \tag{9.12}$$

pentru $s = 0$

$$\mu_0 = (-1)\sigma^2 \cdot \mu_{-2} = 1$$
$$\mu_1 = 0 \cdot \sigma^2 \cdot \mu_{-1} = 0$$
$$\mu_2 = \sigma^2 \cdot \mu_0 = \sigma^2$$
$$\mu_3 = 2 \cdot \sigma^2 \cdot \mu_1 = 0$$

$$\mu_4 = 3 \cdot \sigma^2 \cdot \mu_2 = 3\sigma^2 \cdot \sigma^2 = 3\sigma^4 \tag{9.13}$$

.....

Scanned with OKEN Scanner

În baza relaţiei (9.13), relaţia (9.11) devine:

$$\beta_2 = \frac{\mu_4}{\sigma^4} = \frac{3\sigma^4}{\sigma^4} = 3 \qquad (9.14)$$

Relaţia (9.14) arată că în cazul unei distribuţii normale (mezocurtice) $\beta_2 = 3$, când $\beta_2 > 3$, distribuţia este leptocurtică, iar dacă $\beta_2 < 3$ distribuţia este platicurtică.

Coeficientul de boltire β_2 se întâlneşte în unele lucrări de specialitate şi într-o formă modificată prin scăderea constantei 3, în aşa fel încât γ_2 să fie nul în cazul unei distribuţii normale:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 \qquad \text{sau} \qquad \gamma_2 = \beta_2 - 3 \qquad (9.15)$$

Atât β_2 cât şi γ_2 sunt indicatori fără dimensiune, invariabili prin schimbarea originii şi a unităţii de măsură a caracteristicii şi se interpretează astfel:

$\gamma_2 > 0$ pentru distribuţii leptocurtice;

$\gamma_2 < 0$ în cazul distribuţilor platicurtice;

$\gamma_2 = 0$ în cazul distribuţiilor mezocurtice (normale).

Aplicând acest coeficient în cazul distribuţiei din tabelul 9.1, rezultă:

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{5217{,}57}{37{,}782} - 3 = 1{,}076$$

$\gamma_2 = 1{,}076$ înseamnă că distribuţia este leptocurtică, are loc o concentrare importantă a mamelor de vârstă apropiată de medie.

Scanned with OKEN Scanner

IV. TEORIA SI PRACTICA SONDAJELOR

Scanned with OKEN Scanner

10. SONDAJELE STATISTICE

11. VERIFICAREA IPOTEZELOR STATISTICE

Scanned with OKEN Scanner

10. SONDAJELE STATISTICE

10.1. Noţiunea, importanţa şi felurile sondajelor

10.1.1. Noţiuni fundamentale

Cunoaşterea trăsăturilor specifice ale unor colectivităţi, a distribuţiilor unităţilor lor după diferite caracteristici, se realizează în condiţii obişnuite pe baza cercetării valorilor caracteristicilor tuturor unităţilor lor componente, în mod exhaustiv. Gradul de cunoaştere va fi afectat într-o anumită măsură de erori, dacă se va încerca să se caracterizeze o colectivitate pe baza studierii numai a unei anumite *părţi reprezentative*, extrase din cadrul ei, presupunând că această parte posedă trăsături esenţiale apropiate de cele ale colectivităţii generale. Aceste erori provin din faptul că subcolectivitatea extrasă, reprezintă prin definiţie, numai o parte a colectivităţii generale şi de aceea ea nu poate să aibă exact aceleaşi trăsături caracteristice ca şi colectivitatea din care provine. Totuşi, practica statistică arată că, în anumite condiţii, erorile ce se produc nu sunt de mari proporţii.

Ipoteza potrivit căreia o subcolectivitate poate să reprezinte în mod adecvat, dar incomplet, întreaga colectivitate din care a fost extrasă, se bazează pe o fundamentare logică. Din practica statistică s-a constatat că, cu cât volumul subcolectivităţii este mai mare, cu atât ea va reprezenta mai bine trăsăturile colectivităţii generale, iar în cazul limită, când volumul subcolectivităţii coincide cu cel al colectivităţii din care a fost extrasă, evident, reprezentativitatea va fi perfectă.

Cercetările prin sondaj, urmăresc tocmai să determine în ce măsură caracterizarea unei colectivităţi se poate efectua generalizând proprietăţile unei subcolectivităţi reprezentative, extrase din ea. De exemplu, se poate aprecia media şi abaterea standard a unei colectivităţi cu ajutorul mediei şi abaterii standard obţinute pe baza unei subcolectivităţi reprezentative. Desigur, această problemă poate fi extinsă şi asupra altor indicatori ai colectivităţilor.

De exemplu, cumpărătorul de cereale din toate timpurile, pentru a alege marfă de o anumită calitate recurgea, şi o mai face şi astăzi la fel, la o examinare selectivă. Astfel, pe baza analizei mai mult sau mai puţin amănunţite a unei cantităţi mici de boabe luate dintr-un sac îşi forma o imagine asupra conţinutului calitativ al întregului sac. În mod analog, pentru a analiza un balot dintr-o anumită ţesătură se recurge la studierea unei bucăţi sub aspectul fineţii ţesăturii, modului de imprimare a culorilor, rezistenţei, greutăţii pe unitate de suprafaţă etc. Medicul, pentru precizarea

Scanned with OKEN Scanner

unui diagnostic, recurge adesea la cercetări prin sondaj a unei mici părţi din substanţele componente ale corpului uman (analiza sângelui, analiza unor ţesuturi în laborator etc.).

Este clar că, în exemple ca cele de mai sus, nu trebuie să se acorde o importanţă deosebită alegerii propriu-zise a eşantionului, deoarece materialele considerate sunt omogene prin natura lor, în aşa fel încât se poate lua proba din oricare parte a lor şi aceasta va fi reprezentativă.

Problema alegerii subcolectivităţii n se complică în cazul unor colectivităţi neomogene, când o singură probă nu va mai satisface cerinţa de a reprezenta întreaga colectivitate N. Alegerea va fi mai puţin reprezentativă dacă se vor selecta probele, în mod deliberat, cu scopul de a exclude anumite categorii de unităţi.

Deşi se utilizează de mult timp în cercetarea statistică, metoda sondajelor a fost ştiinţific fundamentată abia în a doua jumătate a secolului al *XIX*-lea, pe baza calculului probabilităţilor.

Esenţa metodei sondajelor constă în următoarele:

- se dă (sau se cunoaşte) o colectivitate formată dintr-un număr mare de unităţi, notat cu N;
- colectivitatea N trebuie caracterizată statistic, din punct de vedere al caracteristicilor $X, Y, \ldots, Z$ (funcţie de repartiţie, medie, dispersie etc.), reprezentate prin variabilele aleatoare, după cum urmează:

$$X: (x_1, x_2, \ldots, x_i, \ldots, x_N)$$
$$Y: (y_1, y_2, \ldots, y_i, \ldots, y_N)$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Z: (z_1, z_2, \ldots, z_i, \ldots, z_N)$$

- din colectivitatea N, care face obiectul cercetării, se extrage, după anumite principii, o subcolectivitate formată din n unităţi $(n < N)$, care trebuie să reprezinte la scară redusă trăsăturile esenţiale ale întregii colectivităţi N. Pentru fiecare unitate statistică extrasă, se înregistrează caracteristicile $X, Y, \ldots, Z$ obţinându-se:

$$X: (x_1', x_2', \ldots, x_i', \ldots, x_n')$$
$$Y: (y_1', y_2', \ldots, y_i', \ldots, y_n')$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$Z: (z_1', z_2', \ldots, z_i', \ldots, z_n')$$

Scanned with OKEN Scanner

• pe baza studierii detaliate a subcolectivităţii n, se cere să se caracterizeze colectivitatea N sub aspectul diferitelor caracteristici.

Colectivitatea N delimitată în timp, în spaţiu şi din punct de vedere calitativ poartă denumirea de *colectivitate generală*.

Unităţile extrase din cadrul colectivităţii generale poartă denumirea de *unităţi de selecţie*, iar acestea, ca orice unitate statistică, pot fi simple sau complexe.

Subcolectivitatea n se mai numeşte *subcolectivitate de selecţie, mostră, eşantion* sau *probă*.

Colectivitatea generală poate fi formată dintr-un număr finit sau infinit de elemente. De exemplu, colectivitatea satelor dintr-o ţară, colectivitatea clădirilor de locuit dintr-un oraş, numărul întreprinderilor industriale dintr-o ţară, numărul butucilor de viţă de vie dintr-o fermă viticolă, numărul cărţilor dintr-o bibliotecă etc., toate sunt colectivităţi formate din elemente concrete, reale şi finite. Exemple clasice de colectivităţi formate dintr-un număr infinit de unităţi sunt cele referitoare la evenimentele rezultate din aruncarea unui zar sau a unei monede. Numărul ipotetic al experimentelor posibile sugerează colectivităţi infinite.

Majoritatea conceptelor statisticii matematice s-au elaborat pornind de la colectivităţi infinite. În cercetarea statistică concretă, în orice domeniu al realităţii, se aplică raţionamente specifice colectivităţilor infinite, unor colectivităţi de volum mare, dar finite, fără ca prin aceasta să se comită erori importante în cercetare.

În general, cu ajutorul cercetărilor prin sondaj, în cazul unei caracteristici cantitative, se pot realiza următoarele obiective:

• *estimarea unor valori tipice* (media, dispersia sau asimetria, coeficienţi de regresie şi corelaţie etc.) ale colectivităţilor pe baza informaţiilor furnizate de către un eşantion;

• *determinarea preciziei estimaţiilor* făcute pe baza sondajelor.

Teoria şi practica statisticii furnizează un mare număr de metode de estimare a unor valori tipice pe bază de eşantioane. Adoptarea uneia sau alteia dintre metode se face în funcţie de natura colectivităţii studiate, de precizia rezultatelor ce urmează a fi obţinute din cercetare, de mărimea cheltuielilor de cercetare, de durata cercetării şi de multe alte condiţii. Uneori, indicatorii estimaţi prin sondaje, se utilizează pentru comparaţii cu indicatorii colectivităţii generale şi pentru stabilirea cauzelor abaterilor dintre cele două categorii de indicatori (reali şi estimaţi).

Organele de decizie, care dispun efectuarea unor cercetări prin sondaj, ca şi oricare alt utilizator al rezultatelor unor astfel de cercetări, trebuie să nu piardă nici un moment din vedere faptul că aceste rezultate sunt posibile, uneori cu un înalt grad de probabilitate, dar niciodată ele nu reprezintă certitudini. De multe ori, în activitatea umană cotidiană, anumite evenimente sunt posibile într-un grad atât de înalt încât întotdeauna le considerăm ca sigure. Alte evenimente sunt atât de nesigure încât întotdeauna vom ezita să luăm anumite decizii mai importante pe baza lor.

Scanned with OKEN Scanner

De aceea, atât cercetătorii, cât şi organele de decizie îşi pun tot mai des problema: care este gradul de încredere pe care putem să-l acordăm estimaţiilor făcute pe bază de sondaj? Cu alte cuvinte, se cere să se determine, în condiţii probabilistice, care este precizia unei estimaţii, în ce măsură se abate estimaţia de la valoarea reală a unei caracteristici dintr-o colectivitate. Dacă precizia estimaţiei nu se poate exprima în termeni probabilistici, atunci va trebui să se apeleze la intuiţie sau la comparaţii cu unele cercetări anterioare.

Precizia oricărei estimaţii depinde de natura colectivităţii N, de modul de constituire a eşantionului şi de metoda de estimare.

10.1.2. Reprezentativitatea eşantionului

Un eşantion este reprezentativ pentru o anumită colectivitate dacă el prezintă, la scară redusă, aceleaşi trăsături esenţiale ca şi colectivitatea generală din care a fost extras: aceeaşi medie, acelaşi grad de împrăştiere şi aceeaşi structură.

La fel ca orice altă activitate umană, cercetarea prin sondaj este afectată de anumite erori. În mod normal, în cadrul cercetărilor prin sondaj, se produc: *erori de înregistrare*, comune tuturor tipurilor de înregistrare statistică şi *erori de reprezentativitate*, specifice înregistrărilor prin sondaj.

Ca şi în cazul cercetărilor exhaustive, erorile de înregistrare ce se produc cu ocazia unui sondaj sunt determinate de acţiunea unor factori întâmplători, cum ar fi: lipsa de memorie, neatenţia personalului de înregistrare etc. Aceste erori au un caracter întâmplător, nesistematic. Alte erori de înregistrare pot apărea în mod sistematic, având un caracter premeditat sau nepremeditat, ca de exemplu încercarea de a majora sau diminua intenţionat mărimea unor indicatori ce urmează a fi înregistraţi.

Reprezentativitatea unui eşantion se apreciază cu ajutorul erorilor de reprezentativitate. *Eroarea de reprezentativitate* se calculează ca o diferenţă dintre valoarea unui indicator sintetic (medie, dispersie etc.) calculat pe baza unui eşantion şi valoarea aceluiaşi indicator calculat pe baza colectivităţii generale. Pentru eşantioanele "perfect" reprezentative, aceste diferenţe vor fi nule, deci structura eşantionului va fi identică cu structura colectivităţii generale din care provin. Desigur, în cercetările concrete nu ne putem aştepta ca eşantionul să asigure o reprezentativitate perfectă a colectivităţii generale. Reprezentativitatea perfectă este un caz ideal spre care se tinde în cercetare, dar care practic nu poate fi atins.

Erorile de reprezentativitate sunt inevitabile, ele se produc în orice cercetare prin sondaj şi pot fi doar diminuate prin respectarea întocmai a principiilor de cercetare, dar nu pot fi complet eliminate. Pe baza teoriei calculului probabilităţilor, se poate determina mărimea lor probabilă. Erorile de reprezentativitate se produc

Scanned with OKEN Scanner

datorită faptului că nu se asigură tuturor unităţilor statistice din colectivitatea generală aceeaşi probabilitate de a face parte din eşantion.

Convenţional, se admite că un eşantion are o reprezentativitate satisfăcătoare, dacă erorile de reprezentativitate sunt mai mici sau cel mult egale cu 5%:

$$\frac{|\bar{x} - m|}{\bar{x}} \cdot 100 \leq 5\%$$

în care: $\bar{x}$ = media de sondaj;

m = media colectivităţii generale.

După cauzele care determină producerea lor, *erorile de reprezentativitate* pot fi *sistematice* şi *întâmplătoare*.

Erorile sistematice de reprezentativitate sunt determinate de nerespectarea principiilor stabilite pentru constituirea eşantionului, care determină unităţilor din cadrul colectivităţii generale şanse neegale de a fi cuprinse în eşantion.

Erorile întâmplătoare de reprezentativitate sunt determinate de însăşi natura cercetărilor prin sondaj şi se pot estima anticipat în orice tip de sondaj.

Ele se produc chiar în cazul respectării riguroase a principiilor de sondaj, deoarece, în mod întâmplător structura eşantionului poate să difere de structura colectivităţii din care a fost extras. Desigur, pentru a se asigura în ansamblu un eşantion reprezentativ, este necesar ca erorile întâmplătoare să fie cât mai mici. Ordinul de mărime al acestor erori depinde de volumul eşantionului, de gradul de împrăştiere a variantelor diferitelor variabile în cadrul colectivităţii generale, de metoda de sondaj utilizată, ca şi de metodele de calcul şi analiză statistică adoptate.

Pentru reducerea mărimii erorilor întâmplătoare de reprezentativitate, se poate proceda la mărirea volumului eşantionului care, în cazul menţinerii tuturor celorlalte condiţii neschimbate, duce la micşorarea lor în mod corespunzător. În general, se poate afirma că *mărimea erorilor întâmplătoare de reprezentativitate variază invers proporţional cu rădăcina pătrată din volumul eşantionului*. Alteori, pentru a se reduce erorile întâmplătoare de reprezentativitate se recurge la împărţirea (stratificarea) colectivităţii generale în grupe mai omogene, mai ales în cazul când colectivitatea generală prezintă o mare variabilitate din punct de vedere al diferitelor caracteristici. În felul acesta se elimină posibilitatea neincluderii în eşantion a anumitor grupe importante ale colectivităţii generale.

În cercetările statistice longitudinale prin sondaj, cum este de exemplu cercetarea nivelului de trai pe baza anchetelor selective, este necesar să se verifice periodic, reprezentativitatea eşantionului. Pentru aceasta, se calculează indicatorii sintetici, cum sunt media, dispersia şi mărimile relative de structură, pentru principalele caracteristici ale unităţilor statistice cuprinse în sondaj şi care se compară apoi cu

Scanned with OKEN Scanner

indicatorii corespunzători ai colectivităţii generale. Dacă diferenţele sunt mai mici de 5% şi se situează între limitele unui interval probabilistic dinainte stabilit, înseamnă că eşantionul se menţine în continuare reprezentativ, iar în caz contrar trebuie să se refacă colectivitatea de selecţie.

10.1.3. Avantajele sondajelor

Erorile de sondaj cuprind atât erori de înregistrare, cât şi erori de reprezentativitate. *Într-un sondaj erorile de înregistrare sunt mult mai mici* decât în cazul unei cercetări exhaustive a colectivităţilor, deoarece, se ia în studiu numai o parte din unităţile statistice ale colectivităţii generale, iar personalul utilizat în cercetările selective este de regulă de o calificare mai înaltă.

În cercetările statistice prin sondaj, se poate utiliza un *program de cercetare mai amănunţit* decât în cazul cercetărilor totale, obţinându-se informaţii care să caracterizeze sub aspecte multiple colectivitatea luată în studiu, iar în faza analizei şi prelucrării acestora se poate asigura o calitate superioară a rezultatelor, datorită faptului că se lucrează cu cadre specializate.

Cercetarea prin sondaj constituie uneori *singura modalitate posibilă* de studiere a unei anumite colectivităţi. Ne referim mai ales la colectivităţile formate dintr-un număr infinit de unităţi, dar şi la cele al căror volum, deşi finit, este foarte mare, ca şi atunci când, prin cercetare unităţile statistice se distrug. În cazul cercetării statistice a calităţii producţiei din punct de vedere al rezistenţei la rupere, la torsiune, la alungire a fibrelor, ţesăturilor, barelor de metal şi altele, durata medie de ardere a becurilor şi tuburilor fluorescente, evaluarea recoltei probabile în lan, prin cercetare, probele se distrug şi atunci cercetarea prin sondaj se impune de la sine.

Datorită faptului că volumul eşantionului este mult mai mic decât volumul colectivităţilor din care acesta provine, cercetarea prin sondaj, deşi se bazează de regulă pe un program de cercetare mai amănunţit, permite *obţinerea rezultatelor mai operativ* şi cu *cheltuieli mai mici* decât în cazul cercetărilor totale. În fond, scopul principal al unei cercetări prin sondaj este acela de a obţine cu un efort minim, un volum maxim de informaţie cu privire la colectivităţile studiate.

În anumite cercetări statistice exhaustive, pentru a se obţine mai rapid unele rezultate preliminare, se recurge la *prelucrarea selectivă a informaţiilor statistice provenite dintr-o înregistrare totală.*

De exemplu, în cazul unui recensământ, analiza volumului imens de date poate fi mult accelerată, dacă pentru prelucrare vor fi selectate numai o parte dintre formularele de recensământ.

O însemnătate deosebită capătă sondajele în cazul cercetărilor care necesită măsurători şi observări directe, pentru care sunt necesare cunoştinţe profesionale deosebite şi, deci, utilizarea unui personal de înaltă calificare. În aceste cazuri, dacă

Scanned with OKEN Scanner

s-ar proceda la cercetarea tuturor elementelor componente ale unor colectivităţi numeroase, cheltuielile de cercetare ar avea un nivel foarte ridicat. Dar, chiar dacă s-ar putea aloca fondurile necesare, este puţin probabil că, într-o perioadă scurtă de timp, s-ar putea instrui într-o măsură suficientă întregul personal necesar pentru cercetare.

Sondajele statistice sunt utilizate uneori şi ca o *metodă de verificare a exactităţii datelor obţinute printr-o înregistrare exhaustivă.*

10.1.4. Tipuri de sondaje

Teoria şi practica statistică cunoaşte un număr mare de categorii de sondaje şi de aceea, pentru studiul lor sistematic, se impune cu necesitate clasificarea acestora după criterii diferite.

După volumul eşantionului, sondajele se pot împărţi în două categorii: *sondaje de volum mare* şi *sondaje de volum redus*. Sondajele de volum mare presupun luarea în studiu a unui eşantion format din cel puţin 120 de unităţi statistice, iar dacă $n < 30$, avem de a face cu sondaje de volum redus. Primele se utilizează în cercetarea nivelului de trai pe baza bugetelor de familie, în cercetările de opinie, în cercetarea populaţiei umane după diferite caracteristici (fertilitate, migraţie etc), iar cea de a doua categorie îşi găseşte o largă aplicabilitate mai ales în controlul statistic al calităţii producţiei în timpul desfăşurării procesului de producţie.

La rândul lor, sondajele de volum mare, în funcţie de numărul de faze necesare pentru constituirea eşantionului, se împart în *sondaje într-o fază* şi *sondaje în două sau mai multe faze.*

În cazul sondajelor de volum redus ($n < 30$), nu se pot stabili limite suficient de apropiate pentru estimaţiile făcute cu privire la diferiţi indicatori ai colectivităţilor. Pentru acest tip de sondaje, nu se poate pune problema determinării valorii unui indicator al colectivităţii generale pe baza estimaţiei lui dintr-un sondaj, ci se verifică doar dacă indicatorul obţinut prin sondaj este semnificativ pentru colectivitatea din care provine eşantionul. În aceste cazuri se testează numai semnificaţia anumitor estimaţii cu ajutorul aşa numitelor *teste de semnificaţie*. Desigur, testele de semnificaţie se aplică şi în cazul sondajelor de volum mare.

Sondajele de volum mare într-o fază, după procedeul utilizat pentru constituirea eşantionului, se pot împărţi în:

- *sondaje aleatoare (întâmplătoare);*
- *sondaje dirijate;*
- *sondaje mixte.*

Sondajul aleator se obţine prin alegerea în mod întâmplător din cadrul colectivităţii generale a unităţilor ce vor intra în componenţa eşantionului. Prin

Scanned with OKEN Scanner

alegere *întâmplătoare* se înţelege că *fiecare unitate a colectivităţii generale are aceeaşi probabilitate de a fi extrasă*. Acesta reprezintă principiul fundamental al sondajului întâmplător. În mod corespunzător, prin eşantion aleator se va înţelege un eşantion astfel ales încât, în cadrul procesului de sondaj, toate eşantioanele posibile de acelaşi volum să aibă aceeaşi probabilitate de a se realiza. Metodele utilizate pentru a asigura caracterul aleator al selecţiei depind de volumul şi natura colectivităţii luate în studiu. Anumite metode care sunt corespunzătoare pentru sondajele de volum mare vor fi necorespunzătoare în cazul sondajelor de volum redus şi invers.

În aparenţă, constituirea unui eşantion aleator din cadrul unei colectivităţi nu ridică probleme deosebite. În realitate, psihologii demonstrează că fiinţa umană este prin natura ei predispusă la o alegere care nu este pur întâmplătoare, obiectivă. De aceea, ori de câte ori este posibil să se identifice şi să se numeroteze unităţile statistice ale unei colectivităţi, se va recurge la *procedeul tragerii la sorţi* sau la utilizarea *tabelelor cu numere întâmplătoare*.

În condiţiile *procedeului tragerii la sorţi* sau al *loteriei* se întocmeşte câte o fişă de aceeaşi mărime, greutate, formă şi culoare pentru fiecare unitate a colectivităţii generale. Pachetul de fişe va constitui, astfel, o reprezentare miniaturală a colectivităţii generale în vederea sondajului. Pentru a extrage un eşantion de volum n din cadrul colectivităţii generale de volum N, se amestecă bine fişele din pachet şi apoi se extrag întâmplător n fişe.

Dificultăţile ridicate de reprezentarea miniaturală a colectivităţilor şi de amestecare a fişelor lor sunt deosebit de mari în cazul în care volumul acestora este foarte mare, iar pentru colectivităţile infinite, acest tip de sondaj este inaplicabil.

Pentru a se evita neajunsurile metodei de sondaj tip loterie se pot înlocui fişele cu numere întâmplătoare cuprinse în aşa numitele *tabele cu numere întâmplătoare* sau pe baza unui program de generare a numerelor întâmplătoare pe un calculator electronic. Problema principală ce se pune cu acest prilej este verificarea faptului dacă numerele sunt cu adevărat întâmplătoare. Se cunosc şi se utilizează pe scară largă în literatura de specialitate numeroase tabele de numere întâmplătoare.

Ca şi metodele de sondaj tip loterie, cea bazată pe utilizarea tabelelor cu numere întâmplătoare se poate folosi numai pentru selecţii din colectivităţi finite.

În cazul colectivităţilor formate dintr-un număr infinit de unităţi se impune apelarea la tehnici adecvate de sondaj. Să considerăm, de exemplu, că vrem să facem un sondaj aleator dintr-o cisternă de 5000 litri de lapte în scopul cercetării diferitelor caracteristici ale acestuia. Este de la sine înţeles că nu putem cerceta toate moleculele din cisterna de lapte şi nici nu putem, în condiţii obişnuite, să izolăm una dintre ele şi să o analizăm separat. Se poate reduce acest caz la cel al unei colectivităţi finite. Pentru aceasta, împărţim laptele din cisternă în două părţi egale şi ne oprim la întâmplare asupra uneia, neglijând-o pe cealaltă. Apoi, jumătatea luată în considerare o împărţim încă o dată în două şi din nou neglijăm una dintre cele două jumătăţi.

Scanned with OKEN Scanner

Procedeul se aplică în continuare în mod similar, până când "jumătatea" devine acceptabilă ca dimensiune.

Există destul temei pentru a considera că proba de lapte, astfel obţinută, are un caracter întâmplător. În practică însă nu este raţional să se procedeze atât de riguros, deoarece se poate aprecia că cei 5000 de litri de lapte formează o colectivitate cu un grad ridicat de omogenitate şi proba se poate lua din oricare parte a cisternei în mărime convenabilă. Acelaşi procedeu se poate utiliza în cazul cerealelor, făinii, cimentului şi, în general în toate colectivităţile cu caracter de substanţă continuă. In cazul unor colectivităţi ipotetice, procedeele descrise mai sus sunt inaplicabile. Să considerăm un exemplu tipic al colectivităţilor infinite şi anume, cel al evenimentelor rezultate din aruncarea unei monede. Nu putem obţine o imagine miniaturală a colectivităţii cu ajutorul sondajului tip loterie, deoarece nu cunoaştem natura colectivităţii. De asemenea, nu se poate alcătui o grămadă a aruncărilor monedei pe care apoi s-o divizăm în mod succesiv pentru constituirea eşantionului. Singura modalitate de constituire a eşantionului aleator este acum de a consemna rezultatele aruncării monedei de un anumit număr de ori, convenabil, drept un eşantion.

Sondajul dirijat se constituie prin alegerea intenţionată a unităţilor din cadrul unei colectivităţi generale, a cărei structură şi parametri sunt, de regulă, cunoscuţi. Sondajul dirijat se deosebeşte de sondajul aleator prin faptul că, deşi şi acesta se desfăşoară pe baza unui plan prestabilit, lasă loc opiniilor, aprecierilor şi preferinţelor personale ale cercetătorului, încălcând în acest fel principiul alegerii întâmplătoare. Aceasta dă naştere la anumite erori de reprezentativitate. Uneori, în practică, sondajul dirijat poate să ducă la obţinerea unei medii de selecţie mai apropiată de media colectivităţii generale decât sondajul aleator, datorită excluderii conştiente de către cercetător a unităţilor statistice ale căror valori se abat în mare măsură de la grupul mediu.

Spre deosebire de sondajul aleator, în cazul sondajului dirijat nu există metode de a stabili gradul de precizie al unei estimaţii. De asemenea, se ştie că pe măsură ce volumul său se măreşte, eşantionul aleator devine din ce în ce mai reprezentativ, proprietate care nu se mai confirmă în cazul sondajelor dirijate. În general, eşantioanele dirijate oferă o imagine concludentă a mediei, dar cea asupra dispersiei poate fi cu totul necorespunzătoare.

Sondajul dirijat se utilizează în practică în prospectarea pieţii şi în cercetările de opinie, unde interpretarea personală a cercetătorului are un câmp larg de manifestare.

Sondajul mixt îmbină avantajele sondajului aleator cu cele ale sondajului dirijat. Forma principală de sondaj mixt, întâlnită în practica statistică, este cea a sondajului tipic sau stratificat, care se efectuează prin împărţirea colectivităţii generale în grupe omogene (straturi sau clase) şi apoi se extrage din fiecare grupă, după anumite principii câte un eşantion aleator.

Scanned with OKEN Scanner

10.2. Sondaje de volum mare

10.2.1. Sondajul întâmplător simplu

Sondajul întâmplător simplu este un caz particular al sondajului aleator de unităţi statistice simple, care se caracterizează prin următoarele trăsături specifice:

- fiecare unitate a colectivităţii generale are aceeaşi probabilitate de a face parte din eşantion;
- extragerea fiecărei unităţi din cadrul colectivităţii generale este independentă de oricare alta;
- volumul colectivităţii generale rămâne acelaşi pe parcursul tuturor extragerilor de constituire a eşantionului.

În practica statistică, cel mai frecvent se utilizează sondajele întâmplătoare extrase din colectivităţi finite.

Principiile de constituire a eşantionului în cazul unui sondaj întâmplător se bazează pe *schema probabilistică clasică a urnei lui Bernoulli cu bila revenită sau cea a bilei nerevenite (schema hipergeometrică).*

În funcţie de schema probabilistică ce se utilizează pentru constituirea eşantionului, *sondajul aleator simplu poate fi repetat* sau *nerepetat.*

10.2.1.1. Sondajul întâmplător simplu repetat

Sondajul simplu repetat are loc atunci când fiecare unitate a colectivităţii generale N, care este extrasă, se reintroduce în cadrul colectivităţii, după ce a fost cercetată, în aşa fel încât volumul colectivităţii generale rămâne neschimbat pe toată durata eşantionării.

Acest tip de sondaj se fundamentează ştiinţific pe schema probabilistică a urnei lui *Bernoulli* cu bila revenită, cu care, de fapt, este în corespondenţă biunivocă.

Dintr-o colectivitate generală de volum N, se poate extrage un număr de $\frac{N^n}{n!}$ sondaje simple repetate de volum n, în aşa fel încât două eşantioane oarecare să difere între ele cel puţin prin natura unei singure unităţi de sondaj. Desigur, dacă N este infinit, atunci şi numărul de sondaje simple repetate de volum n este infinit.

Scanned with OKEN Scanner

În ambele cazuri, sondajele posibile sunt echiprobabile, iar variabilele de selecţie sunt independente.

Pentru prezentarea unora dintre problemele ce se întâlnesc cu ocazia estimării mediei necunoscute a unei colectivităţi, pe baza unui eşantion de volum mare, în cazul unui sondaj întâmplător, considerăm colectivitatea mamelor din *România* distribuite pe grupe cincinale de vârstă (tatăl făcând parte din grupa de vârstă 35-39 ani) în anul 1975 (tabelul 10.1).

Deoarece intervalele de la începutul şi sfârşitul seriei sunt deschise, pentru a evita unele calcule aproximative ce se impun în asemenea cazuri, vom neglija cele două intervale în continuare. Acest fapt nu va avea o influenţă importantă asupra rezultatelor cercetării, deoarece, intervalele marginale reprezintă, de fapt, abateri de la dezvoltarea normală a fenomenului (în cazul nostru, vârsta fertilă normală exclude cele două intervale marginale excepţionale) şi, în plus, ele au o frecvenţă foarte mică în comparaţie cu celelalte intervale. Se observă, din tabelul 10.1 (col.4) că distribuţia născuţilor-vii se apropie în mare măsură de o distribuţie normală.

Tabelul 10.1. Distribuţia mamelor care au născut în anul 1975

Vârsta mamei	Numărul mamelor (tatăl de 35-39 ani)	n_i	f_i	N_i
sub 15 ani	2			
15 - 19 ani	323	323	0,6	323
20 - 24 ani	3868	3868	7,8	4191
25 - 29 ani	13639	13639	27,4	17830
30 - 34 ani	18629	18629	37,4	36459
35 - 39 ani	11861	11861	23,8	48320
40 - 44 ani	1443	1443	2,9	49763
45 - 49 ani	45	45	0,1	49808
50 ani +	2			
Total	49812	49808	100,0	-

Sursa: *Prelucrat după Anuarul Statistic al României, 1976, DCS, Bucureşti, p.28-29*

Principalii indicatori ai colectivităţii generale sunt: $m = 31{,}75$ ani; $\sigma^2 = 24{,}56$ şi $\sigma = 4{,}956$ ani.

Din colectivitatea mamelor N = 49808, se pot extrage întâmplător $\frac{49808^{100}}{100!}$ de eşantioane de volum n = 100, pentru a servi la estimarea unor valori tipice. De regulă, aceşti indicatori vor fi diferiţi de la un eşantion la altul, dar dacă numărul eşantioanelor este suficient de mare, atunci valorile tipice, ce se vor obţine,

Scanned with OKEN Scanner

vor forma o distribuţie de frecvenţe, care se va apropia din ce în ce mai mult de modelul unei curbe continue, de repartiţie teoretică cunoscută.

Pentru ilustrarea acestei idei se vor extrage, pe baza tabelelor cu numere întâmplătoare, 125 de eşantioane de volum $n = 100$ mame din colectivitatea prezentată în tabelul 10.1. Colectivitatea generală fiind formată din 49808 unităţi statistice, presupunem că fiecărei unităţi i se atribuie un număr de ordine de la 1 la 49808.

Alegerea unui eşantion de volum $n = 100$ mame echivalează cu alegerea a 100 de numere întâmplătoare cuprinse între 1 şi 49808. Se aleg, aşadar, primele 100 de numere întâmplătoare mai mici decât 49808, începând din oricare punct al tabelului cu numere întâmplătoare formate din cinci cifre. Astfel, să presupunem că primul număr întâmplător cuprins între 1 şi 49808, întâlnit în tabel, este 5001. Pentru a vedea ce vârstă are mama care a născut acest copil, se determină frecvenţele cumulate N_i (col.5, tabel 10.1). Se observă că, acesta se include în frecvenţa cumulată corespunzătoare intervalului al treilea (25-29 ani). Următorul număr întâmplător pe coloană este 51237, care este mai mare decât 49808, deci nu va fi luat în considerare ş.a.m.d., până se completează eşantionul de volum $n = 100$. Repetând acelaşi procedeu pentru 125 de puncte diferite ale tabelului cu numere întâmplătoare se vor obţine 125 de eşantioane.

Distribuţiile mamelor după vârstă, pe baza celor 125 de eşantioane, se prezintă în tabelul 10.2. Pe ultimele trei coloane ale acestui tabel s-au determinat mediile, dispersiile şi abaterile standard.

Calculând media mediilor de selecţie $\overline{\overline{x}} = 31{,}704$ ani, vom observa că aceasta se apropie în mare măsură de media reală a colectivităţii $\bar{x}$. Mediile de sondaj sunt cuprinse între 30,85 ani (în sondajul 48) şi 32,60 (în sondajul 97), deci foarte aproape de media reală a colectivităţii.

Repartiţia mediilor de selecţie, atunci când numărul eşantioanelor dintr-o colectivitate este suficient de mare, tinde către o funcţie continuă:

$$y = f(x)$$

în care y reprezintă frecvenţa, iar x, variantele variabilei.

Se observă că cele două distribuţii nu sunt pe deplin regulate, deoarece frecvenţa totală, dată de numărul eşantioanelor obţinute din colectivitatea generală, este relativ mică.

În cazul sondajelor de volum mare, media de selecţie poate fi considerată ca o estimaţie a mediei colectivităţii generale.

Scanned with OKEN Scanner

Tabelul 10.2. Distribuţia mamelor pe vârste în colectivitatea generală şi în sondaje

Colecti-vitatea generală	Grupa de vârstă a mamei(ani)		15-19	20-24	25-29	30-34	35-39	40-44	45-49	Total	Media $\bar{x}$	Dispersia s^2	Abaterea standard s
	Născuţi - vii (n_i)		323	3868	13639	18629	11861	1443	45	49808	31,75	24,56	4,96
0	1	2	3	4	5	6	7	8	9	10	11	12	13
Sondaje	Distribuţia născuţilor-vii în sondajul nr.	1	2	8	26	44	18	2	-	100	31,20	23,81	4,88
		2	-	10	24	35	29	2	-	100	31,95	24,95	4,99
		3	-	9	25	35	28	3	-	100	32,05	25,00	5,00
		4	-	11	28	33	25	3	-	100	31,55	26,35	5,13
		5	1	10	24	34	28	3	-	100	31,85	27,83	5,28
		6	1	11	26	32	28	2	-	100	31,55	27,85	5,28
		7	1	9	25	31	33	1	-	100	31,95	26,45	5,14
		8	1	8	27	28	34	2	-	100	32,10	27,34	5,23
		9	-	6	34	29	29	2	-	100	31,85	23,33	4,83
		10	-	7	33	31	27	2	-	100	31,70	23,36	4,83
		11	-	10	31	36	20	3	-	100	31,25	24,19	4,92
		12	2	8	22	45	20	2	1	100	31,65	26,53	5,15
		13	2	7	21	46	22	1	1	100	31,80	25,01	5,00
		14	2	4	23	42	26	2	1	100	32,30	24,96	4,99
		15	1	7	24	42	24	1	1	100	31,90	24,14	4,91
		16	1	5	34	38	21	1	-	100	31,30	20,56	4,53
		17	-	8	36	35	18	2	1	100	31,15	23,93	4,89
		18	-	10	34	34	20	1	1	100	31,05	24,65	4,96
		19	-	10	34	32	22	2	-	100	31,10	24,04	4,90
		20	-	12	28	31	26	2	1	100	31,55	29,00	5,39

Scanned with OKEN Scanner

Tabelul10.2.Continuare

0	1	2	3	4	5	6	7	8	9	10	11	12	13
		21	-	11	25	37	25	2	-	100	31,60	24,69	4,97
		22	-	12	25	34	26	3	-	100	31,65	27,03	5,20
		23	1	11	26	32	27	3	-	100	31,60	28,69	5,36
		24	1	8	29	30	32	-	-	100	31,70	24,86	4,99
		25	1	11	26	31	28	3	-	100	32,00	28,25	5,31
		26	1	9	28	27	34	1	-	100	31,85	27,33	5,23
		27	1	9	25	44	18	2	1	100	31,45	25,15	5,01
		28	-	7	34	29	28	2	-	100	31,70	23,86	4,88
		29	1	9	30	25	34	1	-	100	31,75	27,69	5,26
		30	-	8	34	27	29	2	-	100	31,65	25,03	5,00
		31	-	10	25	35	29	1	-	100	31,80	24,01	4,90
		32	-	9	25	34	29	3	-	100	32,10	25,34	5,03
		33	-	11	25	34	27	1	2	100	31,90	29,14	5,40
		34	1	10	25	34	26	2	2	100	31,90	31,14	5,58
		35	1	11	26	32	27	1	2	100	31,70	31,36	5,60
		36	1	8	27	30	33	1	-	100	31,95	25,95	5,09
		37	1	7	27	28	35	2	-	100	32,25	26,69	5,16
		38	-	7	33	30	28	2	-	100	31,75	23,69	4,87
		39	-	7	35	30	27	1	-	100	31,50	22,50	4,74
		40	-	10	30	37	21	2	-	100	31,25	23,19	4,82
		41	1	9	27	43	18	2	-	100	31,20	22,81	4,78
		42	2	8	22	46	19	2	1	100	31,60	26,19	5,12
		43	2	7	21	48	20	1	1	100	31,70	24,36	4,94

Scanned with OKEN Scanner

Tabelul 10.2. Continuare

0	1	2	3	4	5	6	7	8	9	10	11	12	13
		44	2	4	22	44	26	1	1	100	32,25	23,69	4,87
		45	1	6	25	42	24	1	1	100	31,95	23,45	4,84
		46	-	5	36	35	22	1	1	100	31,53	23,11	4,81
		47	-	8	35	35	19	2	1	100	31,25	24,19	4,92
		48	-	10	35	36	17	1	1	100	30,85	23,53	4,85
		49	-	9	35	33	21	1	1	100	31,15	24,43	4,94
		50	-	12	25	33	27	2	1	100	31,75	28,69	5,36
		51	-	9	34	33	21	3	-	100	31,25	24,18	4,92
		52	1	4	25	40	27	3	-	100	32,35	22,21	4,71
		53	1	7	32	38	21	1	-	100	31,20	21,81	4,67
		54	-	5	38	38	18	1	-	100	31,10	18,04	4,25
		55	1	8	34	34	21	2	-	100	31,10	24,24	4,92
		56	1	7	29	37	20	6	-	100	3,80	27,01	5,20
		57	-	6	26	42	20	4	2	100	32,30	25,96	5,09
		58	1	6	27	34	30	2	-	100	32,10	24,34	4,93
		59	1	7	24	40	28	-	-	100	31,85	21,82	4,67
		60	-	6	26	47	18	3	-	100	31,80	19,76	4,45
		61	-	5	30	39	24	2	-	100	31,90	20,14	4,49
		62	-	4	44	32	15	5	-	100	31,15	21,93	4,68
		63	-	6	31	30	31	2	-	100	32,10	23,34	4,83
		64	1	3	29	36	25	6	-	100	32,45	25,07	5,01
		65	-	9	27	42	18	2	2	100	31,65	26,03	5,10
		66	1	4	28	32	30	5	-	100	32,55	25,75	5,07
		67	-	5	34	31	28	2	-	100	31,90	22,14	4,71

Scanned with OKEN Scanner

Tabelul 10.2. Continuare

0	1	2	3	4	5	6	7	8	9	10	11	12	13
		68	1	5	24	42	25	3	-	100	32,20	22,41	4,73
		69	-	7	30	36	26	1	-	100	31,70	21,36	4,62
		70	-	4	43	31	17	5	-	100	31,30	22,56	4,75
		71	-	3	30	35	30	2	-	100	32,40	19,99	4,47
		72	-	8	32	41	17	2	-	100	31,15	20,42	4,52
		73	-	11	31	34	19	4	1	100	31,35	28,43	5,33
		74	-	6	32	38	22	2	-	100	31,60	20,69	4,55
		75	1	5	28	34	30	2	-	100	32,15	23,62	4,86
		76	2	6	29	37	20	6	-	100	31,75	28,36	5,32
		77	-	6	26	41	23	4	-	100	32,15	22,13	4,70
		78	1	8	21	36	32	2	-	100	32,30	25,46	5,05
		79	-	3	38	36	16	7	-	100	31,80	23,01	4,80
		80	2	5	32	36	23	2	-	100	31,40	24,15	4,91
		81	1	6	28	36	23	6	-	100	32,10	26,84	5,18
		82	1	7	26	39	22	4	1	100	32,00	27,24	5,22
		83	-	10	26	34	24	6	-	100	32,00	22,25	5,32
		84	-	6	34	30	26	4	-	100	31,90	24,64	4,96
		85	2	5	33	35	23	2	-	100	31,40	24,29	4,93
		86	1	6	27	37	23	6	-	100	32,15	26,62	5,16
		87	1	7	26	39	22	4	1	100	32,00	27,25	5,22
		88	1	4	30	44	19	2	-	100	31,60	19,69	4,44
		89	2	7	28	33	25	4	1	100	31,90	30,64	5,54
		90	-	2	37	40	18	3	-	100	31,65	18,03	4,25
		91	-	7	35	29	25	4	-	100	31,70	25,36	5,04

Scanned with OKEN Scanner

Tabelul 10.2. Continuare

0	1	2	3	4	5	6	7	8	9	10	11	12	13
		92	2	6	32	34	24	2	-	100	31,40	25,29	5,03
		93	-	8	29	32	28	2	1	100	32,00	26,25	5,12
		94	1	4	29	32	28	6	-	100	32,50	26,50	5,10
		95	-	7	33	31	28	1	-	100	31,65	22,53	4,75
		96	-	7	41	28	19	5	-	100	31,20	25,31	5,03
		97	1	6	20	40	30	2	1	100	32,60	24,73	4,97
		98	-	6	28	47	19	-	-	100	31,45	16,65	4,08
		99	2	6	27	40	19	6	-	100	31,80	27,51	5,24
		100	-	8	28	42	18	2	2	100	31,70	25,36	5,04
		101	1	5	27	34	29	4	-	100	32,35	25,23	5,02
		102	-	6	40	25	28	1	-	100	31,40	22,90	4,77
		103	-	9	35	33	20	3	-	100	31,15	23,94	4,89
		104	1	5	25	41	25	3	-	100	22,50	22,75	4,77
		105	-	7	27	43	23	-	-	100	31,60	22,38	4,73
		106	-	4	37	37	20	2	-	100	31,45	19,15	4,38
		107	-	5	35	39	17	4	-	100	31,50	21,00	4,58
		108	-	11	21	42	24	2	-	100	31,75	23,69	4,87
		109	-	5	33	36	23	3	-	100	31,80	21,51	4,64
		110	-	5	43	29	20	3	-	100	31,15	21,93	4,68
		111	1	6	28	37	26	1	1	100	31,53	24,78	4,98
		112	-	8	29	38	23	2	-	100	31,60	22,19	4,71
		113	-	7	33	42	18	-	-	100	31,05	17,65	4,20
		114	-	5	30	38	25	2	-	100	31,95	20,45	4,52
		115	-	4	43	31	19	2	1	100	31,25	22,19	4,71
		116	1	5	30	37	25	1	1	100	31,85	23,83	4,88

Scanned with OKEN Scanner

Tabelul 10.2. Continuare

0	1	2	3	4	5	6	7	8	9	10	11	12	13
		117	-	10	23	40	25	2	-	100	31,80	23,51	4,85
		118	-	6	34	40	18	2	-	100	31,30	19,56	4,42
		119	-	5	31	34	27	3	-	100	32,10	22,34	4,73
		120	-	11	21	41	25	2	-	100	31,80	24,01	4,90
		121	-	4	39	34	18	5	-	100	31,55	22,55	4,73
		122	-	5	43	32	20	2	-	100	31,05	20,65	4,54
		123	1	5	32	35	25	1	1	100	31,75	24,19	4,92
		124	-	5	30	37	24	4	-	100	32,10	18,43	4,29
		125	-	8	31	40	19	2	-	100	31,30	21,06	4,59

Scanned with OKEN Scanner

Tabelul 10.3. Distribuţia mediilor de sondaj

Vârsta mamei (ani)	Numărul eşantioanelor
- 30,90	1
30,91 - 31,10	6
31,11 - 31,30	19
31,31 - 31,50	10
31,51 - 31,70	27
31,71 - 31,90	27
31,91 - 32,10	18
32,11 - 32,30	9
32,31 - 32,50	6
32,51 -	2
Total	125

Tabelul 10.4. Distribuţia abaterilor standard

Abaterea standard	Numărul eşantioanelor
- 4,20	2
4,21 - 4,40	4
4,41 - 4,60	12
4,61 - 4,80	22
4,81 - 5,00	42
5,01 - 5,20	24
5,21 - 5,40	16
5,41 -	3
Total	125

În estimarea dispersiei unei colectivităţi, cu ajutorul dispersiei obţinute pe baza unui eşantion, se pot comite anumite erori de estimaţie, deoarece abaterile variantelor variabilei x se calculează în funcţie de media de selecţie şi nu faţă de media reală a colectivităţii. De aceea, pentru a se diminua importanţa acestor eventuale erori, formula obişnuită a dispersiei se modifică prin luarea în considerare la numitorul ei a numărului gradelor de libertate, devenind:

- pentru o distribuţie cu frecvenţă:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_s)^2 n_i}{\sum_{i=1}^{k} n_i - 1} \qquad (10.1)$$

- pentru o distribuţie simplă:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_s)^2}{n-1}$$

Scanned with OKEN Scanner

Dacă n este suficient de mare, scăderea unei unităţi de la numitorul obişnuit al formulei dispersiei are o influenţă nesemnificativă asupra indicatorului şi se poate neglija.

Pentru ilustrarea estimării mediei colectivităţii generale, pe baza unui sondaj întâmplător simplu, considerăm sondajul 1 din tabelul 10.2. Media eşantionului este $\bar{x} = 31{,}2$ ani şi, în absenţa oricărei alte informaţii, vom considera că aceasta este o estimaţie a vârstei medii reale m a mamelor.

Aplicând formula (10.1) la sondajul 1 din exemplul anterior, rezultă dispersia de sondaj:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i - 1} = \frac{(17{,}5-31{,}2)^2 \cdot 2 + (22{,}5-31{,}2)^2 \cdot 8 + (27{,}5-31{,}2)^2 \cdot 26 +}{100-1}$$

$$\frac{+(32{,}5-31{,}2)^2 \cdot 44 + (37{,}5-31{,}2)^2 \cdot 18 + (42{,}5-31{,}2)^2 \cdot 2}{100-1} = \frac{2381}{99} = 24{,}05$$

iar abaterea medie pătratică de sondaj va fi: $s = \sqrt{s^2} = \sqrt{24{,}05} = 4{,}904$ ani.

În scopul de a confirma ipoteza de mai sus, estimaţia punctuală trebuie întotdeauna să fie însoţită de informaţii care să ofere posibilitatea de a-i aprecia calităţile. Precizăm că mărimea eşantionului pe care se bazează estimaţia este $n = 100$ şi că abaterea medie pătratică de sondaj (sau eroarea medie de selecţie, cum se mai numeşte) este $s = 4{,}9$ ani.

Abaterea standard a mediei eşantioanelor posibile dintr-o colectivitate statistică, normal distribuită, depinde de abaterea standard a distribuţiei din colectivitatea generală. De aceea, ori de câte ori nu se cunoaşte abaterea standard din cadrul colectivităţii generale, aceasta se poate înlocui cu abaterea standard obţinută pe baza unui eşantion.

Se remarcă, de asemenea, faptul că, în cazul sondajelor simple de volum mare, repartiţiile de selecţie se apropie de modelul unor repartiţii normale şi aceasta, chiar şi atunci când repartiţia în cadrul colectivităţii generale nu este normală.

Scanned with OKEN Scanner

După cum se arată în literatura de specialitate, dacă se ştie că repartiţia în cadrul colectivităţii originare nu prezintă o asimetrie pronunţată, nu mai este nevoie să-i cunoaştem forma exactă.

Dacă într-o cercetare concretă se cunoaşte media colectivităţii generale m, dar nu se cunoaşte şi dispersia ei, atunci estimarea dispersiei colectivităţii generale se poate face pe baza formulei:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - m)n_i}{\sum_{i=1}^{k} n_i} \qquad (10.2)$$

în care: m - media colectivităţii generale;

n_i - frecvenţele distribuţiei de sondaj.

Vom ilustra aplicarea practică a formulei (10.2) pe baza datelor din sondajul nr.1:

$$s^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i - 1} = \frac{(17{,}5-31{,}2)^2 \cdot 2 + (22{,}5-31{,}2)^2 \cdot 8 + (27{,}5-31{,}2)^2 \cdot 26 +}{100-1}$$

$$\frac{+(32{,}5-31{,}2)^2 \cdot 44 + (37{,}5-31{,}2)^2 \cdot 18 + (42{,}5-31{,}2)^2 \cdot 2}{100-1} = \frac{2381}{99} = 24{,}05$$

iar $s = \sqrt{24{,}05} = 4{,}91$ ani.

S-a arătat că mediile aritmetice ale tuturor sondajelor posibile dintr-o colectivitate, formează o variabilă aleatoare, cu media egală cu valoarea medie a colectivităţii generale şi dispersia $\frac{\sigma^2}{n}$:

Scanned with OKEN Scanner

$$s_{\bar{x}}^2 = \frac{\sigma^2}{n} \tag{10.3}$$

iar eroarea medie de sondaj:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \tag{10.4}$$

Aceasta înseamnă că dispersia mediilor de selecţie variază direct proporţional cu dispersia colectivităţii generale şi invers proporţional cu volumul eşantionului.

Eroarea medie pătratică se poate calcula după formula (10.4), dacă se cunoaşte abaterea medie pătratică σ din cadrul colectivităţii generale.

În practica cercetărilor prin sondaj însă, în cele mai multe cazuri, abaterea medie pătratică a colectivităţii generale nu se cunoaşte şi atunci se utilizează estimaţia acesteia pe baza formulei (10.2).

Substituind pe (10.1) în (10.4) rezultă:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x}_s)^2 n_i}{\sum_{i=1}^{k} n_i - 1}} \Big/ n \tag{10.5}$$

În exemplul anterior, eroarea medie pătratică de sondaj pe baza primului eşantion va fi:

$$s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4{,}91}{10} = 0{,}491 \text{ ani} \qquad \text{sau} \qquad s_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4{,}904}{10} = 0{,}4904 \text{ ani}$$

Se poate afirma, cu o anumită probabilitate $P = 2\Phi(t)$, că media colectivităţii generale se va situa în intervalul:

$$(\bar{x} - t \cdot s_{\bar{x}}) < m < (\bar{x} + t \cdot s_{\bar{x}}) \tag{10.6}$$

Scanned with OKEN Scanner

Intervalul (10.6) poartă denumirea de *interval de încredere*, în condiţiile unui anumit nivel de semnificaţie: $\rho = 1 - p$.

Factorul t reprezintă o variabilă aleatoare normală, având valoarea medie egală cu zero şi eroarea medie pătratică egală cu 1 şi este o funcţie de probabilitatea cerută.

Produsul $t \cdot s_{\bar{x}}$ poartă denumirea de *eroare limită de sondaj* sau *eroare maximă admisibilă*:

$$\Delta_x = t \cdot s_{\bar{x}} = \frac{t\sigma}{\sqrt{n}} \qquad (10.7)$$

Din relaţia (10.7) rezultă:

$$t = \frac{\Delta_x}{s_{\bar{x}}}$$

Cu cât creşte probabilitatea cu care vrem să garantăm rezultatele, cu atât creşte şi valoarea lui t şi, în mod corespunzător, mărimea intervalului în care se va încadra media colectivităţii generale. Astfel, pentru:

$t = 1{,}0$	$\Phi = 0{,}3413$	iar	$P(1{,}0) = 2\Phi(t) = 0{,}6826$
$t = 1{,}5$	$\Phi = 0{,}4332$		$P(1{,}5) = 0{,}8664$
$t = 2{,}0$	$\Phi = 0{,}4772$		$P(2{,}0) = 0{,}9544$
$t = 2{,}5$	$\Phi = 0{,}4938$		$P(2{,}5) = 0{,}9876$
$t = 3{,}0$	$\Phi = 0{,}4987$		$P(3{,}0) = 0{,}9974$
$t = 3{,}5$	$\Phi = 0{,}4998$		$P(3{,}5) = 0{,}9996$
$t = 3{,}7$	$\Phi = 0{,}4999$		$P(3{,}7) = 0{,}9998$

În exemplul nostru, la un nivel de semnificaţie (risc) de $\rho = 0{,}05$ sau probabilitatea de 95% cu care se garantează că media aritmetică a colectivităţii generale va fi cuprinsă în intervalul dat de relaţia (10.6), va fi:

$$P[(\bar{x} - t \cdot s_{\bar{x}}) < m < (\bar{x} + t \cdot s_{\bar{x}})] = 2\Phi(t) \qquad (10.8)$$

pentru un risc de 5% corespunde $2\Phi(t) = 0{,}95$, iar din tabelele cu repartiţia normală normată, rezultă $t = 1{,}96$.

Scanned with OKEN Scanner

Substituind pe t cu valoarea lui particulară în relaţia (10.8) şi pe σ, cu valoarea lui din sondajul 1, se obţine:

$$P[(31{,}2 - 1{,}96 \cdot 0{,}491) < m < (31{,}2 + 1{,}96 \cdot 0{,}491)] = 95\%$$

sau

$$P(30{,}23 < m < 32{,}16) = 95\%$$

Deci probabilitatea ca m să fie cuprins între 30,23 ani şi 32,16 ani este egală cu 95%.

Aceasta înseamnă că în 95 de eşantioane de volum n = 100, din fiecare 100 eşantioane media aritmetică va fi cuprinsă în intervalul [30,23 ani; 32,16 ani] şi numai în 5 eşantioane din 100, media se va situa în afara acestor limite. Mai concret, putem afirma că vârsta medie a mamelor, care au născut în anul 1975 copii ai căror taţi erau de vârstă 35-39 ani, a fost cuprinsă între 30,23 ani şi 32,16 ani şi aceasta se poate garanta cu o probabilitate de 95%.

În cazul în care vom garanta rezultatele cu o probabilitate de $P = 0{,}9898 \approx 0{,}99$, media colectivităţilor generale se va situa în intervalul:

$$\bar{x} \pm 2{,}56 \cdot 0{,}491 = 31{,}2 \pm 1{,}26$$

deci

$$29{,}94 < m < 32{,}46$$

Variaţia în paralel a nivelului de semnificaţie, a probabilităţii de garantare a rezultatelor şi a mărimii intervalului de încredere pentru media colectivităţii generale din exemplul nostru, se prezintă în tabelul 10.5.

Tabelul 10.5. Nivelul de semnificaţie şi mărimea intervalului de încredere

Nivelul de semnificaţie (%)	Probabilitatea $P = 2\Phi(t)$	t	Intervalul $\bar{x} \pm t \cdot s_{\bar{x}}$	Mărimea intervalului d_i (ani)
31,74	68,26	1,0	[30,71; 31,69]	0,98
13,36	86,64	1,5	[30,46; 31,94]	1,48
4,56	95,44	2,0	[30,22; 32,18]	1,96
1,24	98,76	2,5	[29,97; 32,43]	2,46
0,26	99,74	3,0	[29,72; 32,67]	2,95
0,04	99,96	3,5	[29,48; 32,92]	3,44
0,02	99,98	3,7	[29,38; 33,02]	3,63

Scanned with OKEN Scanner

Se observă că, pe măsură ce creşte probabilitatea cu care se garantează rezultatele şi scade în mod corespunzător nivelul de semnificaţie (fiind probabilităţi complementare), variabila aleatoare t se măreşte, iar mărimea intervalului de încredere creşte în mod proporţional.

Sondajul întâmplător repetat se poate efectua şi în cazul unei caracteristici alternative dintr-o colectivitate, formată din elemente A şi elemente *non A*. Să considerăm că extragerea unui element A reprezintă un eveniment favorabil, iar a unui element *non A*, un eveniment nefavorabil. Dacă se extrag L eşantioane aleatoare din colectivitatea formată din N elemente, frecvenţele teoretice ale eşantioanelor cu 0,1,2,3, ... evenimente favorabile sunt date de repartiţia $L \cdot (p+q)^n$. Dezvoltând acest binom, rezultă:

$$L \cdot (p+q)^n =$$

$$= L \cdot [q^n + nq^{n-1}p + \frac{n(n-1)}{2} q^{n-2}p^2 + \cdots + nqp^{n-1} + p^n] \qquad (10.9)$$

unde: p = probabilitatea unui eveniment favorabil " A";
q = probabilitatea unui eveniment nefavorabil " *non A*";

iar $$p + q = 1$$

Repartiţia binomială (10.9) este caracterizată de următoarele valori tipice:

- *media* M, reprezintă numărul absolut de evenimente favorabile în cadrul eşantionului:

$$M = np \qquad (10.10)$$

- *dispersia* $$s_{np}^2 = npq \qquad (10.11)$$

- *eroarea standard* a numărului absolut de evenimente favorabile:

$$s_{np} = \sqrt{npq} \qquad (10.12)$$

Dacă într-un sondaj vom considera, în locul numărului absolut al evenimentelor favorabile, ponderea acestora în cadrul eşantionului $p = \frac{M}{n}$, repartiţia binomială se va caracteriza prin alte valori tipice de selecţie şi anume:

Scanned with OKEN Scanner

- media va fi egală cu p;
- eroarea standard a lui p va fi: $s_p = \sqrt{\frac{pq}{n}}$ (10.13)

Cu ajutorul relaţiilor (10.12) şi (10.13), se poate aprecia dacă diferenţa dintre abaterile standard ale valorilor obţinute dintr-un sondaj şi cele aşteptate se datorează fluctuaţiilor de selecţie sau unor cauze reale.

Ca şi în cazul caracteristicilor nealternative, s_p poartă denumirea de *eroare medie de sondaj*. Pentru valori mari ale eşantionului, repartiţia binomială unimodală (10.9) se apropie de o repartiţie normală, ceea ce determină ca în majoritatea sondajelor întâmplătoare numărul absolut al evenimentelor favorabile să se situeze în intervalul:

$$(np - t \cdot s_p, np + t \cdot s_p)$$

Eroarea maximă admisibilă în cadrul acestui tip de sondaj va fi egală cu:

$$\Delta_{np} = t \cdot s_{np} \quad \text{sau} \quad \Delta_{np} = t\sqrt{npq} \tag{10.14}$$

de unde rezultă:

$$t = \frac{\Delta_{np}}{s_{np}} \tag{10.15}$$

Dacă în relaţia (10.14) substituim pe s_{np} cu s_p, rezultă:

$$\Delta_p = t \cdot s_p = t\sqrt{\frac{pq}{n}} \tag{10.16}$$

sau

$$\Delta_p = t\sqrt{\frac{p(1-p)}{n}} \tag{10.17}$$

În practica sondajelor, dacă nu se cunosc indicatorii p şi q din cadrul colectivităţii generale şi nu se pot calcula pe bază de sondaj sau dacă se pot calcula, dar

Scanned with OKEN Scanner

nu prezintă suficientă încredere, atunci se poate admite nivelul maxim al erorii standard, care are loc când $p = q = \frac{1}{2}$, adică:

$$s_p = \sqrt{pq} = \sqrt{\frac{1}{2} \cdot \frac{1}{2}} = \sqrt{0,25} = 0,5$$

Mai mult, dacă una dintre proporţiile p şi q este foarte mică, atunci

$$pq = p(1-p) = p - p^2 \cong p$$

În aceste condiţii, eroarea standard (10.12) devine:

$$s_{np} = \sqrt{np} \tag{10.18}$$

Relaţia (10.18) arată că, dacă proporţia evenimentelor favorabile este mică, abaterea standard a numărului evenimentelor favorabile este egală (aproximativ) cu rădăcina pătrată a mediei numărului absolut de cazuri favorabile.

10.2.1.2. Sondajul întâmplător simplu nerepetat

Sondajul întâmplător nerepetat are loc atunci când unitatea statistică, extrasă din cadrul colectivităţii N şi cercetată, nu mai revine în colectivitate. Aceasta înseamnă că, pe măsură ce se constituie eşantionul, volumul colectivităţii generale se diminuează cu unităţile selectate. Astfel, după prima extragere, colectivitatea generală va fi compusă din $N-1$ unităţi, după a doua extragere, din $N-2$ şi aşa mai departe, încât după extracţia n vor rămâne în cadrul colectivităţii generale $N-n$ unităţi. Probabilitatea de selecţie creşte pe măsura constituirii eşantionului, astfel, pentru prima unitate este $p_1 = \frac{1}{N}$, pentru a doua, $p_2 = \frac{1}{N-1}$, ... , iar pentru extracţia n va fi $p_n = \frac{1}{N-n+1}$. Schema probabilistică după care se efectuează sondajul simplu nerepetat este cea a urnei lui *Bernoulli* cu bila nerevenită. Sondajul simplu nerepetat nu are sens decât în cazul colectivităţilor formate dintr-un număr finit de elemente.

Scanned with OKEN Scanner

Dintr-o colectivitate de volum N se pot extrage C_N^n eşantioane distincte de volum n. Sondajele sunt echiprobabile, iar variabilele de selecţie sunt dependente.

În statistica matematică se demonstrează că eroarea medie în cadrul sondajului nerepetat este mai mică decât în cazul sondajului repetat:

$$s_{\bar{x}}^2(nerepetat) = s_{\bar{x}}^2(repetat) \cdot \frac{N-n}{N-1}$$

Deoarece, practic, N este un număr foarte mare, se poate renunţa la scăderea unei unităţi de la numitor, fără să se modifice semnificativ valoarea raportului. Eroarea medie de sondaj va deveni:

$$s_{\bar{x}}^2(nerepetat) = \left(1 - \frac{n}{N}\right) s_{\bar{x}}^2(repetat) \qquad (10.19)$$

Substituind pe $s_{\bar{x}}^2(repetat)$ cu expresia sa din (10.3), rezultă:

$$s_{\bar{x}}^2(nerepetat) = \frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right) \qquad (10.20)$$

Eroarea medie de selecţie pentru un sondaj aleator simplu nerepetat va fi:

$$s_{\bar{x}} = \sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (10.21)$$

Dacă nu se cunoaşte dispersia colectivităţii generale, se poate utiliza în locul ei dispersia obţinută pe baza unui eşantion $s_{\bar{x}}^2$.

Eroarea maximă de sondaj va avea expresia:

$$\Delta_x = t s_{\bar{x}} = t\sqrt{\frac{\sigma^2}{n}\left(1 - \frac{n}{N}\right)} \qquad (10.22)$$

Intervalul în care se va cuprinde media colectivităţii generale, garantat cu o probabilitate $P = 2\Phi(t)$, va fi:

Scanned with OKEN Scanner

$$P\left[\bar{x}-t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}<m<\bar{x}+t\sqrt{\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)}\right]=2\Phi(t) \qquad (10.23)$$

În mod analog, eroarea medie de sondaj în cadrul unui sondaj simplu nerepetat, pentru o caracteristică alternativă , va fi:

$$s_p=\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} \quad \text{şi} \quad s_{np}=\sqrt{npq\left(1-\frac{n}{N}\right)} \qquad (10.24)$$

iar eroarea limită:

$$\Delta_p=t\sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} \quad \text{şi} \quad \Delta_{np}=t\sqrt{npq\left(1-\frac{n}{N}\right)} \qquad (10.25)$$

10.2.1.3. Sondajul mecanic

Sondajul mecanic, numit şi *sondaj sistematic*, este un caz particular de sondaj simplu care nu respectă în totalitate cerinţele unui sondaj aleator.

Sondajul mecanic constă în alcătuirea unei liste cu unităţile colectivităţii generale, independent de caracteristicile lor principale şi apoi, pe baza unei fracţii de sondaj, se extrag unităţile ce vor intra în componenţa eşantionului. Întrucât volumul eşantionului n se cunoaşte sau se stabileşte anticipat, fracţia de sondaj d se determină prin raportarea volumului colectivităţii generale la volumul eşantionului:

$$d=\frac{N}{n} \qquad (10.26)$$

Apoi, din primele d unităţi ale listei, se extrage întâmplător o unitate, să presupunem cu numărul de ordine k (în care $1\le k\le d$); unitatea următoare, ce va intra în cadrul eşantionului, va avea numărul de ordine $k+d$, a treia unitate va fi $k+2d$ şi aşa mai departe, iar ultima va avea numărul de ordine $k+(n-1)d$.

De exemplu, să presupunem că, dintr-o colectivitate generală de volum $N=10000$ elemente, se efectuează un sondaj mecanic de volum n = 100, iar fracţia de sondaj d = 100. Cu ajutorul unui procedeu de sondaj întâmplător (prin tragerea la sorţi sau pe baza tabelelor cu numere întâmplătoare), din primele 100 unităţi statistice

Scanned with OKEN Scanner

se extrage una, să presupunem cea cu numărul de ordine 36. Unităţile care vor intra în componenţa eşantionului, vor fi cele cu numerele de ordine: 36,136, 236, 336, ..., 9936.

În cazul sondajului mecanic nu se asigură tuturor unităţilor colectivităţii generale aceeaşi şansă de a face parte din eşantion. Astfel, prima extragere se efectuează numai din o mică parte a unităţilor colectivităţii generale, iar celelalte extrageri depind de numărul de ordine al extrageriilor anterioare. Acest fapt a determinat pe unii statisticieni să denumească acest procedeu de selecţie *sondaj cvasi-aleator*.

Sondajul mecanic se utilizează în practică în cazul unor colectivităţi relativ omogene, peste tot acolo unde unităţile colectivităţii generale sunt (sau pot fi) înscrise în anumite liste, independent de caracteristicile care interesează cercetarea. De exemplu, lista populaţiei în ordine alfabetică, listă existentă la primăriile localităţilor, care cuprind populaţia pe străzi şi numere etc. De asemenea, acest tip de sondaj se poate utiliza şi în agricultură pentru evaluarea recoltei în lan.

Precizia unui sondaj mecanic depinde în mod hotărâtor de natura colectivităţii generale. Astfel, dacă elementele unei colectivităţi, din listă, sunt afectate de o anumită periodicitate, variaţie ciclică sau zonală, rezultatele sondajului sistematic vor fi cu totul nesatisfăcătoare.

Literatura de specialitate recomandă, în cazul sondajului mecanic, utilizarea formulelor de calcul ale erorilor de sondaj stabilite pentru sondajul întâmplător repetat. Folosind aceste formule, se măreşte precizia sondajului, întrucât erorile ce se produc în cazul unui sondaj mecanic sunt, în general, mai mici decât cele ale sondajului repetat şi mai mari decât cele ale sondajului nerepetat.

10.2.1.4. Dimensionarea eşantioanelor întâmplătoare

O problemă deosebit de importantă care apare în faza de pregătire a cercetărilor prin sondaj, este stabilirea *volumului eşantionului*. Erorile de sondaj (medie şi maximă) variază invers proporţional cu rădăcina pătrată din volumul eşantionului. La stabilirea volumului eşantionului, este necesar să se ţină seama de precizia pe care trebuie să o aibă rezultatele sondajului, de probabilitatea cu care trebuie garantate aceste rezultate, de urgenţa cu care trebuie efectuată cercetarea şi de volumul cheltuielilor necesare pentru cercetare.

Legătura dintre volumul eşantionului, pe de o parte, şi probabilitatea cu care se garantează rezultatele, abaterea standard a colectivităţii generale şi eroarea maximă admisibilă în cercetare, pe de altă parte, este dată de formula erorii limită. De aceea, stabilirea formulei volumului eşantionului în cazul unui sondaj întâmplător se poate face pe baza erorii limită.

Scanned with OKEN Scanner

În cazul unui sondaj aleator repetat, corespunzător unei caracteristici nealternative, eroarea limită este dată de formula (10.3). Din această formulă, prin transformări elementare, rezultă imediat volumul eşantionului:

$$\Delta_x = t\frac{s_{\bar{x}}}{\sqrt{n}} \qquad \text{şi} \qquad \Delta_x^2 = \frac{t^2 s_{\bar{x}}^2}{n}$$

de unde
$$n = \frac{t^2 s_{\bar{x}}^2}{\Delta_x^2} \tag{10.27}$$

În mod analog se stabileşte formula volumului eşantionului în cazul unei caracteristici alternative:

$$\Delta_p = t\sqrt{\frac{p(1-p)}{n}}$$

de unde
$$n = \frac{t^2 p(1-p)}{\Delta_p^2} \tag{10.28}$$

În cazul sondajului aleator nerepetat, volumul eşantionului se determină astfel:

- pentru o caracteristică nealternativă:

$$\Delta_x = t\sqrt{\frac{s_{\bar{x}}^2}{n}\left(1-\frac{n}{N}\right)} \qquad \text{iar} \qquad \Delta_x^2 = \frac{t^2 s_{\bar{x}}^2}{n} - \frac{t^2 s_{\bar{x}}^2}{N}$$

de unde
$$n = \frac{t^2 s_{\bar{x}}^2}{\Delta_x^2 + \frac{t^2 s_{\bar{x}}^2}{N}} \tag{10.29}$$

- pentru o caracteristică alternativă:

$$\Delta_p = t\sqrt{\frac{p(1-p)}{n}\left(1-\frac{n}{N}\right)} \qquad \text{iar} \qquad \Delta_p^2 = \frac{t^2 p(1-p)}{n} - \frac{t^2 p(1-p)}{N}$$

Scanned with OKEN Scanner

de unde

$$n = \frac{t^2 p(1-p)}{\Delta_p^2 + \frac{t^2 p(1-p)}{N}} \qquad (10.30)$$

10.2.2. Sondajul de serii

Sondajul de serii reprezintă o variantă particulară a sondajului aleator. Am văzut că sondajul simplu se utilizează în cazul colectivităţilor generale formate din unităţi statistice simple. Pe lângă acest tip de colectiviţăţi, în practica social-economică, ca şi în alte domenii, se întâlnesc numeroase colectivităţi în care unităţile simple se prezintă împărţite pe grupe cu grade diferite de omogenitate, care se pot considera *unităţi complexe*. De exemplu, grupele de studenţi în învăţământul superior, echipele sau brigăzile de lucru din industrie, construcţii, agricultură, colectivitatea clădirilor de locuit dintr-o anumită localitate (formate din locuinţe), colectivitatea familiilor dintr-o populaţie la un moment dat etc. În cazul cercetării selective a unor astfel de colectivităţi, eşantionul se va constitui prin alegerea întâmplătoare a unităţilor complexe. Unităţile complexe mai poartă denumirea de *serii de unităţi simple*, de aceea sondajul, utilizat în acest caz, se numeşte *sondaj de serii*. Pentru fiecare serie, din cadrul eşantionului, nivelul caracteristicii este dat de nivelul mediu al caracteristicii unităţilor simple care o compun. Acest fapt determină necesitatea înregistrării tuturor unităţilor simple ce intră în componenţa unei serii, pentru a se asigura mediei seriei $\bar{x}_j$ un caracter reprezentativ. Nivelul mediu al caracteristicii din cadrul întregului eşantion de serii se va determina ca o medie aritmetică ponderată a mediilor de serie.

Să presupunem că se studiază selectiv o colectivitate generală formată din R serii, din care se extrage un eşantion compus din r serii ($r < R$), tabelul 10.6.

Media caracteristicii dintr-o serie $\bar{x}_j$ se determină după formula:

$$\bar{x}_j = \frac{\sum_{i=1}^{k} x_{ij}}{k} \qquad i = \overline{1,k} \qquad j = \overline{1,r} \qquad (10.31)$$

iar media caracteristicii la nivelul eşantionului $\bar{x}$ va fi calculată ca o medie din medii:

Scanned with OKEN Scanner

$$\bar{x} = \frac{\sum_{i=1}^{r} \bar{x}_j}{r} \qquad j = \overline{1,r} \tag{10.32}$$

dispersia medie de selecţie se va calcula ca o dispersie dintre serii (grupe), astfel:

$$\delta^2 = \frac{\sum_{j=1}^{r} (\bar{x}_j - \bar{x})^2}{r} \tag{10.33}$$

Tabelul 10.6. Distribuţia unităţilor statistice pe variante ale caracteristicii X şi pe serii

X \ Seria	1	2	3	...	j	...	r
x_1	x_{11}	x_{12}	x_{13}	...	x_{1j}	...	x_{1r}
x_2	x_{21}	x_{22}	x_{23}	...	x_{2j}	...	x_{2r}
⋮			⋮				
x_i	x_{i1}	x_{i2}	x_{i3}	...	x_{ij}	...	x_{ir}
⋮			⋮				
x_k	x_{k1}	x_{k2}	x_{k3}	...	x_{kj}	...	x_{kr}

Formulele mediei şi dispersiei de selecţie (10.32) şi (10.33) se utilizează numai în cazul eşantioanelor ale căror serii sunt formate din acelaşi număr k de unităţi simple. Dacă seriile sunt formate dintr-un număr diferit de unităţi, atunci se vor utiliza formulele ponderate ale mediei şi dispersiei.

În statistica matematică se demonstrează că erorile de selecţie sunt mai mari în cazul unui sondaj de serii, decât în cazul unui sondaj simplu individual.

Ca şi în cazul sondajului simplu individual, sondajul de serii poate fi repetat sau nerepetat, în funcţie de schema probabilistică utilizată pentru constituirea eşantionului.

Eroarea medie de selecţie în cazul sondajului de serii va fi:

a) pentru un sondaj de serii repetat:

- în cazul unei caracteristici nealternative:

Scanned with OKEN Scanner

$$s_{\bar{x}} = \sqrt{\frac{\delta^2}{r}} \qquad (10.34)$$

- în cazul unei caracteristici alternative:

$$s_p = \sqrt{\frac{\delta_p^2}{r}} \qquad 10.35)$$

b) pentru un sondaj simplu de serii nerepetat, eroarea medie de selecţie va fi mai mică decât în cazul sondajului de serii repetat şi se determină astfel:

- în cazul unei caracteristici nealternative:

$$s_{\bar{x}} = \sqrt{\frac{\delta^2}{r}\left(1-\frac{r}{R}\right)} \qquad (10.36)$$

- în cazul unei caracteristici alternative:

$$s_p = \sqrt{\frac{\delta_p^2}{r}\left(1-\frac{r}{R}\right)} \qquad (10.37)$$

Coeficientul, care ne permite trecerea de la formulele corespunzătoare erorii medii în cazul sondajului simplu repetat, la formulele corespunzătoare pentru sondajul simplu de serii nerepetat $\frac{R-r}{R-1}$, se modifică datorită renunţării la scăderea unei unităţi din R de la numitor, deoarece acesta nu influenţează esenţial rezultatele, când R este suficient de mare: $\frac{R-r}{R-1} \cong 1-\frac{r}{R}$.

Precizia unui sondaj de serii este influenţată şi de gradul de împrăştiere din cadrul seriilor şi anume, este cu atât mai mare, cu cât împrăştierea din interiorul acestora este mai mică şi invers.

Eroarea maximă de selecţie se va obţine în fiecare caz, pe baza aplicării formulei generale:

Scanned with OKEN Scanner

$$\Delta = t \cdot s_{\bar{x}} \tag{10.38}$$

Formulele de calcul ale erorilor maxime admisibile, pentru diferite tipuri de sondaj de serii, se prezintă în tabelul 10.7.

Tabelul 10.7. Eroarea maximă admisibilă în cazul sondajelor de serii

Felul caracteristicii / Tipul de sondaj	Nealternativă	Alternativă
Repetat	$\Delta_x = t\sqrt{\frac{\delta^2}{r}}$	$\Delta_p = t\sqrt{\frac{\delta_p^2}{r}}$
Nerepetat	$\Delta_x = t\sqrt{\frac{\delta^2}{r}\left(1-\frac{r}{R}\right)}$	$\Delta_p = t\sqrt{\frac{\delta_p^2}{r}\left(1-\frac{r}{R}\right)}$

Pe baza formulelor erorii maxime admisibile din tabelul 10.7, se poate determina volumul eşantionului r în cazul sondajelor aleatoare de serii (tabelul 10.8).

În practică, sondajul de serii poate fi folosit singur sau combinat cu sondajul simplu individual. De exemplu, în prima fază se selectează un anumit număr de serii dintr-o colectivitate şi apoi în faza următoare, din cadrul fiecărei serii se extrage întâmplător un anumit număr de unităţi simple. Erorile de selecţie în cazul unui astfel de sondaj, vor fi mai mari datorită faptului că în fiecare fază s-a comis o anumită eroare, iar eroarea de selecţie totală va trebui să ţină seama de erorile rezultate din toate fazele de constituire a eşantionului.

Tabelul 10.8. Volumul eşantionului în cazul sondajelor aleatoare de serii

Felul caracteristicii / Tipul de sondaj	Nealternativă	Alternativă
Repetat	$r = \frac{t^2\delta^2}{\Delta_x^2}$	$r = \frac{t^2\delta_p^2}{\Delta_p^2}$
Nerepetat	$r = \frac{t^2\delta^2}{\Delta_x^2 + \frac{t^2\delta^2}{R}}$	$r = \frac{t^2\delta^2}{\Delta_p^2 + \frac{t^2\delta_p^2}{R}}$

Scanned with OKEN Scanner

10.2.3. Sondajul tipic

Sondajul tipic constituie o formă principală a sondajului mixt. El *îmbină* unele proprietăţi ale *sondajului aleator* cu cele ale *sondajului dirijat*, fapt care contribuie la îmbunătăţirea reprezentativităţii eşantionului. Astfel, în vederea efectuării unui sondaj tipic, colectivitatea generală se împarte în mai multe grupe "tipice", numite şi straturi, formate dintr-un număr egal sau inegal de unităţi. Gruparea se poate face după una sau mai multe caracteristici. Prin gruparea tipică se urmăresc două obiective principale: mărirea preciziei estimaţiilor diferiţilor indicatori ai colectivităţii generale şi asigurarea unei reprezentări corespunzătoare în eşantion a diferitelor grupe importante din cadrul acesteia.

Rezultatele cercetării prin sondaj vor avea o precizie maximă în cazul în care grupele colectivităţii prezintă un grad ridicat de omogenitate. Dacă, în unele domenii de cercetare, gruparea după o anumită caracteristică nu duce la obţinerea unor grupe omogene, pentru sondaj este necesar ca grupele respective să fie împărţite în grupe mai mici de unităţi cu un grad mai ridicat de omogenitate.

Gruparea tipică influenţează mărimea erorilor de sondaj. Întrucât erorile de selecţie în cazul sondajului tipic sunt determinate de gradul de omogenitate al grupelor colectivităţii, volumul eşantionului trebuie să fie suficient de mare, încât din fiecare grupă să se poată selecta cel puţin 30 de unităţi.

Din fiecare grupă se va selecta un anumit număr de unităţi pe baza principiilor sondajului aleator. Mai des utilizat în practica statistică este sondajul tipic nerepetat după una sau mai multe caracteristici.

Din punct de vedere formal, o problemă de sondaj tipic se enunţă astfel:

Să considerăm o colectivitate C de volum N, formată din k grupe $C_1, C_2, C_3, \ldots, C_i, \ldots, C_k$ de volum N_i.

Pentru fiecare din cele k grupe, caracteristica X ia următoarele valori:

$$
\begin{aligned}
C_1:&\quad (x_{11}, x_{12}, x_{13}, \ldots, x_{1j}, \ldots, x_{1N_1}) \\
C_2:&\quad (x_{21}, x_{22}, x_{23}, \ldots, x_{2j}, \ldots, x_{2N_2}) \\
&\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad i = \overline{1,k} \\
C_i:&\quad (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{ij}, \ldots, x_{iN_i}) \qquad j = \overline{1,N_i} \\
&\quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
C_k:&\quad (x_{k1}, x_{k2}, x_{k3}, \ldots, x_{kj}, \ldots, x_{kN_k})
\end{aligned}
$$

Scanned with OKEN Scanner

Volumul colectivităţii generale C este egal cu suma volumelor celor k grupe componente:

$$N = N_1 + N_2 + N_3 + \cdots + N_i + \cdots + N_k = \sum_{i=1}^{k} N_i$$

Din fiecare grupă C_i se extrage, după principiile selecţiei aleatoare, câte un număr de unităţi egal cu:

$$n_1, n_2, n_3, \ldots, n_i, \ldots, n_k$$

Volumul eşantionului va fi egal cu:

$$n = n_1 + n_2 + n_3 + \cdots + n_i + \cdots + n_k = \sum_{i=1}^{k} n_i$$

Eşantionul provenit dintr-un sondaj tipic va fi compus din k sondaje aleatoare, corespunzător celor k grupe ale colectivităţii generale. Valorile variabilei X în cele k sondaje vor fi:

$$\begin{aligned}
&C_1': \quad (x'_{11}, x'_{12}, x'_{13}, \ldots, x'_{1j}, \ldots, x'_{1n_1}) \\
&C_2': \quad (x'_{21}, x'_{22}, x'_{23}, \ldots, x'_{2j}, \ldots, x'_{2n_2}) \\
&\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \qquad i = \overline{1,k} \\
&C_i': \quad (x'_{i1}, x'_{i2}, x'_{i3}, \ldots, x'_{ij}, \ldots, x'_{in_i}) \qquad j = \overline{1,n_i} \\
&\qquad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&C_k': \quad (x'_{k1}, x'_{k2}, x'_{k3}, \ldots, x'_{kj}, \ldots, x'_{kn_k})
\end{aligned}$$

Pentru colectivitatea generală, indicatorii tendinţei centrale se vor estima pe baza formulelor:

- media grupei i :

$$m_i = \frac{\sum_{j=1}^{N_i} x_{ij}}{N_i} \qquad i = \overline{1,k} \qquad j = \overline{1,N_i} \qquad (10.39)$$

- media colectivităţii generale:

Scanned with OKEN Scanner

$$m = \frac{\sum_{i=1}^{k} m_i N_i}{\sum_{i=1}^{k} N_i} \quad \text{sau} \quad m = \frac{\sum_{i=1}^{k}\sum_{j=1}^{N_i} x_{ij}}{\sum_{i=1}^{k} N_i} \tag{10.40}$$

În mod analog se vor calcula indicatorii corespunzători ai eşantionului:

- media grupei i :

$$\bar{x}_i = \frac{\sum_{j=1}^{n_i} x'_{ij}}{n_i} \tag{10.41}$$

- media eşantionului:

$$\bar{x} = \frac{\sum_{i=1}^{k} \bar{x}_i n_i}{\sum_{i=1}^{k} n_i} \quad \text{sau} \quad \bar{x} = \frac{\sum_{i=1}^{k}\sum_{j-1}^{n} x'_{ij}}{\sum_{i=1}^{k} n_i} \qquad \begin{matrix} i = \overline{1,k} \\ j = \overline{1,n_i} \end{matrix} \tag{10.42}$$

Dispersiile corespunzătoare colectivităţii generale şi eşantionului vor fi date de formulele:

- pentru colectivitatea generală:

$$\sigma^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{N_i}(x_{ij} - m)^2}{\sum_{i=1}^{k} N_i} \qquad i = \overline{1,k} \qquad j = \overline{1,N_i} \tag{10.43}$$

- pentru eşantion:

Scanned with OKEN Scanner

$$s^2 = \frac{\sum_{i=1}^{k}\sum_{j=1}^{N_i}(x'_{ij} - \bar{x})^2}{\sum_{i=1}^{k} n_i} \qquad i = \overline{1,k} \qquad j = \overline{1,n_i} \qquad (10.44)$$

Problema principală, care trebuie să se rezolve în cazul sondajului tipic, este estimarea valorilor m şi σ^2, pe baza lui $\bar{x}$ şi s^2.

Rezolvarea acestei probleme depinde de modul de stabilire a volumului fiecărei grupe a eşantionului, în raport cu volumul grupei corespunzătoare din cadrul colectivităţii generale. În practică, acest lucru se realizează după principii diferite pe baza cărora se disting trei feluri de sondaje tipice: simple, proporţionale şi optime.

10.2.3.1. Sondajul tipic simplu

Sondajul tipic simplu se caracterizează prin faptul că, din fiecare grupă a colectivităţii generale, se extrage în mod întâmplător un anumit număr de unităţi (acelaşi sau diferit) pentru constituirea eşantionului.

Media sondajului se determină ca o medie aritmetică a mediilor de grupă ponderate cu volumul grupelor după formula (10.45). Dacă grupele din sondaj sunt egale, atunci media sondajului se poate calcula ca o medie aritmetică simplă a mediilor de grupă.

Pentru estimarea mediei colectivităţii generale $\bar{\bar{x}}$ pe baza mediilor de grupă $\bar{x}_i$ din cadrul unui eşantion obţinut printr-un sondaj tipic simplu, se utilizează formula:

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{k}\bar{x}_i N_i}{\sum_{i=1}^{k} N_i} \qquad i = \overline{1,k} \qquad (10.45)$$

În statistica matematică se demonstrează că $\bar{\bar{x}}$ este o estimaţie absolut corectă a mediei caracteristicii din colectivitatea generală, pentru care:

$$M(\bar{\bar{x}}) = m$$

iar

Scanned with OKEN Scanner

$$D^2(\bar{\bar{x}}) = \sum_{i=1}^{k} \frac{s_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)\frac{N_i^2}{N^2} \qquad i = \overline{1,k} \tag{10.46}$$

unde

$$s_i^2 = \frac{\sum_{i=1}^{k}(x_{ij} - \bar{x}_i)^2}{N_i - 1} \tag{10.47}$$

Inegalitatea lui Cebîşev se aplică numai în cazul în care volumul grupelor n_i din eşantion este suficient de mare, iar dispersiile de grupă s_i^2 sunt mici.

Eroarea limită se va determina folosind formula: $\Delta_x = t \cdot s_{\bar{x}}$

Eroarea medie în cazul sondajului tipic simplu se calculează prin extragerea rădăcinii pătrate din dispersia funcţiei de estimaţie:

$$s_{\bar{x}} = \sqrt{\sum_{i=1}^{k} \frac{N_i^2}{N^2}(1 - f_i)\frac{s_i^2}{n_i}} \qquad \text{unde} \qquad f_i = \frac{n_i}{N_i} \tag{10.48}$$

10.2.3.2. Sondajul tipic proporţional

Sondajul tipic proporţional sau reprezentativ se caracterizează prin faptul că, din fiecare grupă a colectivităţii generale, se extrage în mod întâmplător, şi de regulă nerepetat, un număr de unităţi n_i în aşa fel încât raportul dintre volumul unei grupe n_i a eşantionului şi volumul grupei N_i corespunzătoare din colectivitatea generală să fie egal cu raportul dintre volumul total al eşantionului n şi volumul total al colectivităţii N, astfel:

$$\frac{n_i}{N_i} = \frac{n}{N} \qquad i = \overline{1,k} \tag{10.49}$$

adică

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_i}{N_i} = \cdots = \frac{n_k}{N_k} = \frac{\sum_{i=1}^{k} n_i}{\sum_{i=1}^{k} N_i} = \frac{n}{N}$$

Scanned with OKEN Scanner

Din formula (10.49) rezultă că numărul unităţilor statistice care trebuie selectate dintr-o grupă i a colectivităţii generale pentru a forma grupa i a eşantionului se poate stabili cu ajutorul formulei:

$$n_i = \frac{nN_i}{N} \qquad i = \overline{1,k} \tag{10.50}$$

Dispersia funcţiei de estimaţie se determină, ca şi în cazul sondajului tipic simplu, pe baza relaţiei:

$$D^2(\bar{\bar{x}}) = \sum_{i=1}^{k} \frac{s_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)\frac{N_i^2}{N^2} \qquad i = \overline{1,k} \tag{10.51}$$

şi

$$\frac{n_i}{N_i} = \frac{n}{N} = f = const. \tag{10.52}$$

Substituind pe n_i din (10.50) în relaţia (10.51) şi ţinând seama de relaţia (10.52), rezultă:

$$D^2(\bar{\bar{x}}) = \sum_{i=1}^{k} \frac{s_i^2}{\frac{nN_i}{N}}(1-f)\frac{N_i^2}{N^2} \qquad i = \overline{1,k} \tag{10.53}$$

şi

$$D^2(\bar{\bar{x}}) = \frac{1-f}{n}\sum_{i=1}^{k}\frac{s_i^2 N_i}{N} \tag{10.54}$$

Menţionăm faptul că sondajul tipic proporţional este întotdeauna mai eficient decât sondajul tipic simplu, deoarece duce la o dispersie mai mică.

10.2.3.3. Sondajul tipic optim

Numărul de unităţi statistice n_i ce se extrag din grupa i a colectivităţii generale, în cazul unui sondaj tipic proporţional, variază direct proporţional cu volumul total al eşantionului şi cu ponderea grupei i în cadrul colectivităţii generale, dar face abstracţie de gradul de omogenitate din fiecare grupă a colectivităţii N. Dacă pe lângă condiţiile cerute de sondajul tipic proporţional pentru constituirea

Scanned with OKEN Scanner

eşantionului se va lua în consideraţie şi gradul de împrăştiere (abaterea standard) de la nivelul fiecărei grupe a colectivităţii generale, atunci sondajul, astfel obţinut, va avea o eficienţă maximă din punct de vedere al preciziei rezultatelor. Un asemenea sondaj poartă denumirea de *sondaj tipic optim*.

Numărul unităţilor ce urmează a fi selectate din grupa i a colectivităţii generale va fi dat de formula:

$$n_i = \frac{n s_i N_i}{\sum_i s_i N_i} \qquad i = \overline{1,k} \tag{10.55}$$

Dispersia funcţiei de estimaţie se obţine substituind pe n_i din (10.55), în relaţia (10.46), astfel:

$$D^2\left(\bar{\bar{x}}\right) = \sum_{i=1}^{k} \frac{s_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)\frac{N_i^2}{N^2}$$

$$D^2\left(\bar{\bar{x}}\right) = \frac{1}{N^2}\sum_{i=1}^{k}\left(\frac{s_i^2 N_i^2}{n_i} - \frac{s_i^2 n_i N_i^2}{n_i N_i}\right) =$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{k}\frac{s_i^2 N_i^2}{n_i} - \sum_{i=1}^{k} s_i^2 N_i\right] =$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{k}\frac{s_i^2 N_i^2}{\frac{n s_i N_i}{\sum_{i=1}^{k} s_i n_i}} - \sum_{i=1}^{k} s_i^2 N_i\right]$$

Scanned with OKEN Scanner

$$D^2(\bar{\bar{x}}) = \frac{1}{N^2}\left[\frac{\left(\sum_{i=1}^{k} s_i N_i\right)^2}{n} - \sum_{i=1}^{k} s_i^2 N_i\right] \qquad i = \overline{1,k} \qquad (10.56)$$

Menţionăm că sondajul tipic optim este mai eficient decât sondajul tipic proporţional, deoarece dă o dispersie mai mică a mediei de selecţie.

Pe lângă optimizarea din punct de vedere al preciziei unui sondaj se poate asigura şi o minimizare a cheltuielilor totale de sondaj.

Scanned with OKEN Scanner

11. VERIFICAREA IPOTEZELOR STATISTICE

În ultimele decenii, pe plan mondial se acordă o atenţie tot mai mare problemelor de inferenţă statistică, adică celor care se referă la generalizarea rezultatelor obţinute pe baza unui eşantion, în cadrul unei teorii unitare. Cu toate că această teorie are multe avantaje teoretice şi conceptuale, aplicarea ei în practică ridică însă probleme deosebit de dificile. Pentru a înţelege aceste probleme, trebuie să luăm în consideraţie faptul că oricât de obiectiv s-ar planifica o cercetare sau o experienţă, este imposibil să se elimine toate elementele de subiectivism. De exemplu, este cel puţin parţial o decizie subiectivă faptul dacă să ne bazăm un experiment (determinarea unui anumit indicator) pe 5 măsurători, pe 10 măsurători, pe 15 sau mai multe. De asemenea, factorii subiectivi intră inevitabil în proiectarea unui echipament, în retribuirea personalului şi chiar în precizia formulării unei probleme pe care dorim să o investigăm. Un element de subiectivism apare şi atunci când vrem să definim termeni cum ar fi "bun" sau "cel mai bun", în legătură cu alegerea între diferite criterii de decizie (să zicem, când decidem între media unui eşantion şi mediana lui, într-o problemă de estimare), ori când căutăm o linie dreaptă care să corespundă "cel mai bine" unui şir de puncte într-un plan.

În acest capitol vom prezenta numai probleme clasice de inferenţă statistică, astfel încât, din punct de vedere formal, să nu fie luate în consideraţie categoriile separate ale diferiţilor factori subiectivi pe care le-am menţionat mai sus.

Cu alte cuvinte, elementele subiective nu vor apare ca o parte componentă a formulelor, mai degrabă ele vor apare în alegerea diferitelor formule care urmează să fie folosite într-o situaţie dată, în deciziile cu privire la mărimea unui eşantion, în precizarea probabilităţilor cu care vrem să o considerăm acceptabilă şi aşa mai departe. În abordarea bayesiană a problemelor de inferenţă statistică, pot fi luate în calcul o parte din aceste categorii de factori subiectivi.

11.1. Noţiunea de estimare

Potrivit unor dicţionare, o estimaţie este o valoare, o opinie sau aproximaţie făcută pe baza unor date incomplete sau imprecise. Deşi această definiţie se poate aplica şi estimatiilor făcute în mod curent de unii părinţi cu privire la abilităţile (capacităţile) copiilor lor, acestea nu sunt modalităţi de utilizare a termenului de

Scanned with OKEN Scanner

"estimaţie" în statistică. În statistică sunt permise doar estimaţii bazate doar pe informaţii incomplete, numai dacă acestea se sprijină pe un raţionament logic sau pe experienţă şi dacă eşantioanele au fost ştiinţific constituite.

Metodele statistice de estimare îşi găsesc aplicare aproape peste tot în ştiinţă, în activitatea economică, ca şi în viaţa de fiecare zi. În ştiinţele naturii, de exemplu, un biolog vrea să estimeze ce proporţie dintr-o anumită categorie de animale se nasc cu deficienţe fizice; un psiholog doreşte să estimeze timpul mediu necesar unui copil pentru a reacţiona la un stimul dat, un inginer doreşte să estimeze ce variabilitate există în rezistenţa la rupere a unui fir de bumbac etc. În activitatea economică, *Casa de Economii şi Consemnaţiuni* încearcă să estimeze ce proporţie dintre depunătorii săi îşi planifică să-şi cumpere o locuinţă în anul următor, o societate comercială urmăreşte să stabilească salariul mediu al personalului în anul viitor, un cumpărător de lămpi de iluminat doreşte să estimeze variabilitatea duratei lor medii de funcţionare etc.

În sfârşit, în viaţa de toate zilele noi dorim să estimăm care este procentul de accidente de automobil datorate unor defecţiuni de frână, suntem interesaţi să estimăm timpul mediu necesar pentru a călca o cămaşă, să cunoaştem ce variabilitate se poate aştepta în rezultatele la învăţătură ale unui elev, student sau doctorand.

Precizăm că, în fiecare caz, am dat trei exemple: unul referitor la estimarea unui procent sau proporţii, unul cu privire la estimarea unei mărimi medii şi unul referitor la estimarea unei dispersii. Aceştia, de regulă, sunt parametrii pe care-i întâlnim în majoritatea problemelor de estimare.

Revenind la exemplele din ştiinţă, biologul poate să estimeze proporţia animalelor cu deficienţe fizice la 0,1%, psihologul poate să estimeze timpul necesar pentru a reacţiona la un stimul la 0,20 secunde şi inginerul poate să estimeze abaterea standard a măsurătorilor rezistenţei la rupere a diferitelor sortimente de fire la *1kg.f.* Astfel de estimaţii poartă denumirea de *estimaţii punctuale*, deoarece fiecare dintre ele constă dintr-un număr, adică un singur punct de pe axa numerelor reale. Estimaţiile punctuale au serioase neajunsuri, deoarece ele nu ne spun cât de aproape ne putem aştepta să fie faţă de cantităţile pe care se presupune că le estimează. Cu alte cuvinte, *estimaţiile punctuale nu spun nimic despre încrederea sau precizia metodei de estimare care urmează a fi folosită.* Spre exemplu, dacă o reclamă pretinde, pe baza evidenţei "ştiinţifice" că 80 % din toţi fumătorii preferă marca X de ţigări, această concluzie nu va fi semnificativă, dacă se bazează pe opinii luate de la 100 fumători, dintre care s-a întâmplat că 80 preferă marca X. Totuşi, concluzia ar putea deveni mult mai semnificativă dacă ar fi bazată pe interviuri la 1000 de fumători, 5000 de fumători sau poate chiar la 10000 de fumători. Aceasta ilustrează de ce estimaţiile punctuale trebuie întotdeauna să fie însoţite de informaţii care să permită aprecierea calităţii lor.

În majoritatea metodelor discutate în acest capitol, am presupus că estimaţiile noastre sunt bazate numai pe observaţii sau măsurători directe. Dacă acest fel de informaţii urmează să fie suplimentate cu informaţii colaterale, bazate pe date

Scanned with OKEN Scanner

indirecte ori pe raţionamente subiective ale unor persoane, poate fi necesar să folosim forma bayesiană de inferenţă statistică.

11.2. Verificarea ipotezelor

A testa o ipoteză cu privire la un indicator al tendinţei centrale înseamnă a decide dacă se acceptă sau se respinge o ipoteză cu privire la indicatorul corespunzător al unei colectivităţi pe baza informaţiilor obţinute dintr-un eşantion.

O anumită ipoteză poate fi adevărată sau falsă, iar cercetătorul poate să o accepte sau să o respingă. Din combinarea acestor două modalităţi rezultă situaţia din tabelul 11.1.

Tabelul 11.1. Decizie şi eroare în statistică

Decizia \ Ipoteza H	Ipoteza H este adevărată	Ipoteza H este falsă
Acceptă ipoteza H	Decizie corectă	Eroare de genul II
Respinge ipoteza H	Eroare de genul I	Decizie corectă

Decizia va fi corectă dacă se acceptă o ipoteză adevărată sau dacă se respinge o ipoteză falsă. Decizia va fi incorectă în celelalte două situaţii. Dacă se respinge o ipoteză adevărată, se va comite o *eroare de genul I*, iar dacă se acceptă o ipoteză falsă, se va comite o *eroare de genul II*. Eroarea poate fi de genul I sau II, în funcţie de felul în care este formulată ipoteza pe care vrem s-o testăm. Probabilitatea de a respinge ipoteza H când este falsă se numeşte *puterea criteriului*, notată cu $\pi(m)$.

Probabilitatea comiterii erorilor de genul II este egală cu 1 minus puterea criteriului şi se notează cu q_{II}:

$$q_{II} = 1 - \pi(m)$$

Gradul de încredere al oricărei estimaţii a unui indicator statistic, obţinut pe baza unui eşantion, depinde în primul rând de *stabilitatea* acestui indicator de sondaj. Ca măsură a stabilităţii unui indicator în statistică, se folosesc indicatorii împrăştierii. Tocmai aceşti indicatori se utilizează în teoria şi practica statistică drept criterii de apreciere a semnificaţiei estimaţiilor obţinute prin sondaj. De asemenea, estimaţiile

Scanned with OKEN Scanner

vor fi cu atât mai semnificative, cu cât volumul eşantionului va fi mai mare. Acest fapt determină necesitatea luării în calcul a volumului eşantionului în formularea oricărui criteriu de semnificaţie în statistică.

Criteriile de semnificaţie capătă diferite expresii, în funcţie de natura problemei ce trebuie rezolvată.

În formularea unui criteriu de semnificaţie întotdeauna se porneşte de la o ipoteză de lucru, numită în statistică *ipoteza diferenţei nule, notată cu* H_0. Această ipoteză postulează faptul că nu există nici o diferenţă semnificativă între indicatorii obţinuţi pe bază de sondaj şi cei ai colectivităţii generale sau între indicatorii obţinuţi cu ajutorul a două sondaje. Emiterea ipotezelor urmează, în general, regula potrivit căreia *formularea ipotezelor trebuie să se facă într-o astfel de manieră, încât să ştim la ce ne putem aştepta dacă ea se verifică*. Pentru a respecta această regulă, adesea trebuie să presupunem exact opusul a ceea ce vrem să dovedim. De exemplu, dacă vrem să demonstrăm că există o diferenţă esenţială în ce priveşte gradul de precizie în funcţionarea a două maşini similare, vom formula ipoteza că nu există nici o diferenţă din punct de vedere al preciziei de funcţionare; dacă vrem să arătăm că un nou medicament este mai bun decât altul mai vechi, vom formula ipoteza că nu există nici o diferenţă între ele, deci ambele au acelaşi efect; dacă vrem să arătăm că un minereu de fier dintr-o zonă are un conţinut mai bogat în fier decât minereul din altă zonă vom formula ipoteza că ambele minereuri au acelaşi conţinut de fier etc.

Din aceste exemple, rezultă cum a luat naştere noţiunea de ipoteză nulă, cu toate că în prezent ea este folosită pentru orice ipoteză cu scopul de a stabili mai întâi dacă poate fi respinsă. Menţionăm faptul că ideea stabilirii unei ipoteze nule este obişnuită, chiar în afara sferei statisticii. Desigur, acest lucru trebuie să se întâmple la tribunal cu ocazia unui proces, unde acuzatul este considerat nevinovat dacă fapta sa nu este dovedită fără posibilitatea de a fi tăgăduită. Ipoteza că acuzatul este nevinovat este o ipoteză nulă.

Dacă prin aplicarea unui test de semnificaţie, diferenţa dintre valoarea aşteptată şi cea obţinută prin calcule este aşa de mare, încât nu se poate în mod rezonabil atribui întâmplării, ipoteza nulă se respinge. Dacă diferenţa între ceea ce se aşteaptă şi ceea ce s-a obţinut este aşa de mică, încât poate fi atribuită întâmplării, se spune că ea nu este semnificativă din punct de vedere statistic.

Aplicarea unui test de semnificaţie pentru verificarea unei ipoteze statistice se face în următoarele etape:

1. Se formulează o ipoteză (nulă) în aşa fel încât probabilitatea erorii de genul I să poată fi calculată;

2. Se formulează ipoteza alternativă, astfel încât respingerea ipotezei nule să fie echivalentă cu acceptarea ipotezei alternative;

3. Se specifică probabilitatea cu care suntem dispuşi să acceptăm o eroare de genul II pentru valoarea alternativă dată a parametrului pe care îl testăm;

Scanned with OKEN Scanner

5. Se specifică dacă alternativa respingerii ipotezei (1) este acceptarea ei sau reluarea raţionamentului.

11.2.1. Compararea mediei unei colectivităţi cu o valoare dată

Problema testării ipotezei nule, privind egalitatea dintre media unei anumite colectivităţi m şi o anumită valoare dată m_0, înseamnă a verifica dacă între m şi m_0 există o diferenţă suficient de mare care nu poate fi determinată de întâmplare sau dacă aceasta diferenţă este nesemnificativă şi se datorează întâmplării. În mod obişnuit ipoteza alternativă este de forma: $m < m_0$; $m > m_0$ sau $m = m_0$, ceea ce înseamnă că media colectivităţii este mai mică, mai mare sau neegală cu m_0 din cadrul ipotezei nule. Dacă simbolizăm cu H_0 ipoteza nulă şi cu H_A ipoteza alternativă, atunci:

$$H_0 : m = m_0 \tag{11.1}$$

$$H_A : m < m_0, \quad m > m_0, \quad m \neq m_0$$

Criteriul de verificare a ipotezei nule H_0, pe baza mediei de selecţie $\bar{x}$, are următoarea expresie:

$$z = \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}} \tag{11.2}$$

Folosind unităţile standardizate z, criteriul (11.2) devine aplicabil unei mari varietăţi de probleme. Menţionăm că abaterea medie pătratică a colectivităţii σ este presupusă cunoscută.

Dacă vom considera că distribuţia mediilor de selecţie este normală, aplicarea criteriului de verificare a ipotezei nule, în funcţie de alegerea ipotezei nule, se prezintă în figura 11.1.

Scanned with OKEN Scanner

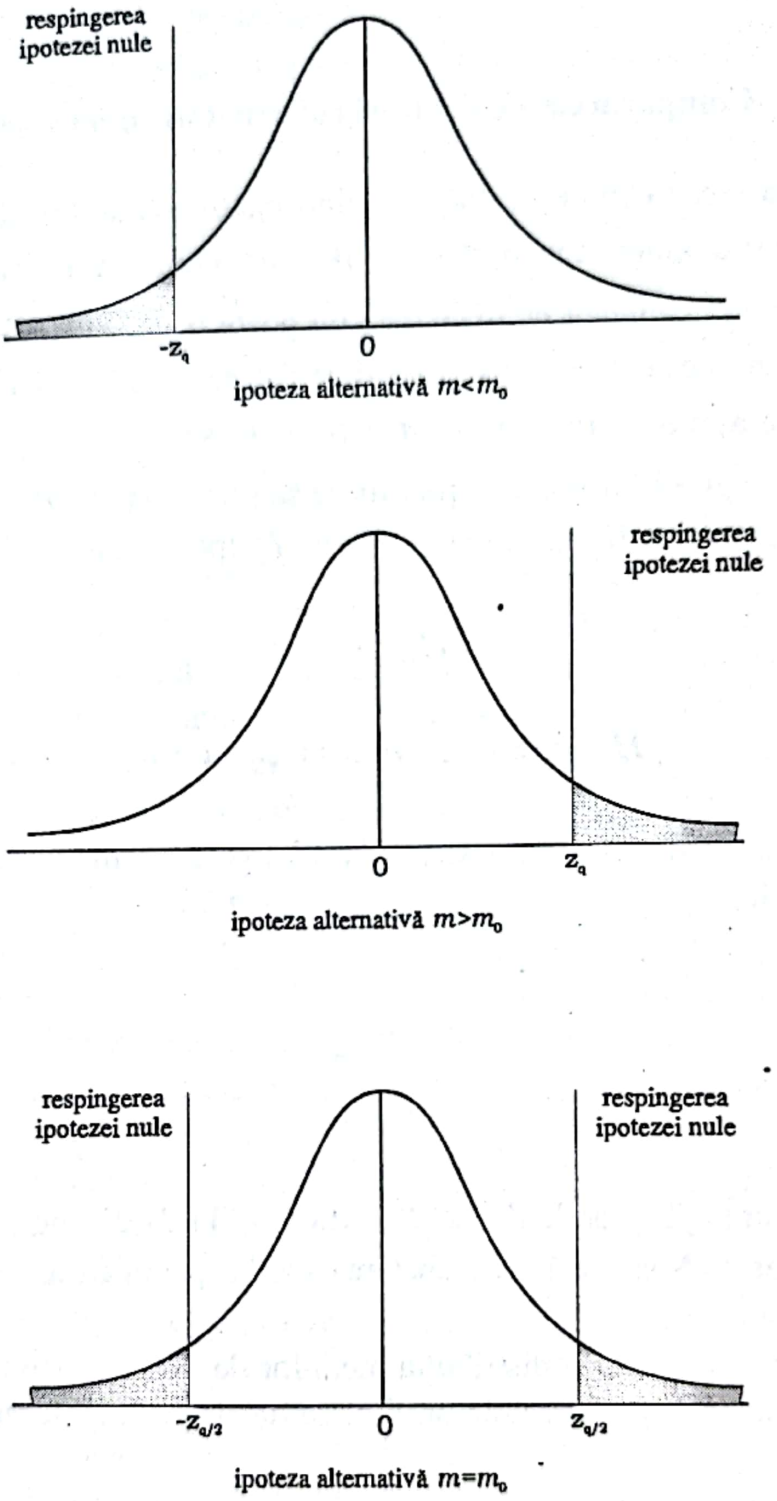


Figura 11.1. Distribuţia normală şi verificarea ipotezei nule

Scanned with OKEN Scanner

Simbolic, criteriul de verificare a ipotezei $H_0: m = m_0$ se prezintă în tabelul 11.2.

Tabelul 11.2. Verificarea ipotezei nule $m = m_0$

Ipoteza alternativă	Respingerea ipotezei nule dacă	Acceptarea ipotezei nule ori continuarea verificării dacă
$m < m_0$	$z < -z_q$	$z \geq -z_q$
$m > m_0$	$z > z_q$	$z \leq z_q$
$m \neq m_0$	$z < -z_{q/2}$ sau $z > z_{q/2}$	$-z_{q/2} \leq z \leq z_{q/2}$

Cele mai des utilizate valori pentru nivelul de semnificaţie sunt $q_1 = 0{,}05$ şi $q_1 = 0{,}01$.

În cazul sondajelor de volum mare ($n > 30$), dacă nu se cunoaşte abaterea medie pătratică a colectivităţii generale, atunci ea se poate înlocui cu cea de sondaj, iar criteriul (11.2) devine:

$$z = \frac{\bar{x} - m_0}{\frac{s}{\sqrt{n}}} \qquad (11.3)$$

Pentru ilustrarea utilizării acestui criteriu de semnificaţie, să considerăm colectivitatea celor 49808 mame, din tabelul 11.2 şi să verificăm, pe baza sondajului 1 de volum $n = 100$ şi $\bar{x}_1 = 31{,}2$ ani, dacă media colectivităţii generale, pe care considerăm că nu o cunoaştem, diferă semnificativ de 32 ani. Presupunem, pentru început, că se cunoaşte abaterea medie pătratică din colectivitatea generală $\sigma = 5{,}96$ ani şi un nivel de semnificaţie $q_1 = 0{,}01$.

Cunoscând $q_1 = 0{,}01$ iar $q_1 = \frac{1}{2} - \Phi(z)$

$$0{,}01 = \frac{1}{2} - \Phi(z)$$

$$\Phi(z) = 0{,}49$$

Scanned with OKEN Scanner

iar din tabelele cu funcţia de repartiţie normală găsim $z_{0,01} = 2,33$.

Substituind datele din problemă în formula (11.2) se obţine:

$$z = \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}} = \frac{31,2 - 32}{4,96\sqrt{100}} = \frac{-0,8}{0,496} = -1,61$$

Deci $z > -z_q$ adică - 1,61 > - 2,33, iar ipoteza alternativă $m < m_0$ se respinge şi se acceptă ipoteza nulă şi anume, media colectivităţii generale nu diferă în mod semnificativ de vârsta de 32 ani şi aceasta în $P = 2\Phi(z) =$ 98 cazuri din 100.

În cazul în care nu am fi cunoscut parametrul σ, acesta s-ar fi putut înlocui cu s calculat pe baza eşantionului 1, astfel:

$$z = \frac{\bar{x} - m_0}{\frac{s}{\sqrt{n}}} = \frac{31,2 - 32}{4,88\sqrt{100}} = -1,64$$

Valoarea lui z, obţinută în acest caz, nu diferă semnificativ de cea obţinută anterior şi decizia nu se va schimba.

Să rezolvăm, în continuare, aceeaşi problemă la un nivel de semnificaţie $q_1 = 0,05$:

$$0,05 = \frac{1}{2} - \Phi(z) \quad \text{de unde} \quad \Phi(z) = 0,45 \quad \text{iar} \quad z_{0,05} = 1,65$$

$$z = \frac{\bar{x} - m_0}{\frac{\sigma}{\sqrt{n}}} = -1,61$$

şi în acest caz ipoteza nulă se acceptă.

La un nivel de semnificaţie $q_1 = 0,06$, vom obţine:

$$\Phi(z) = 0,44 \quad \text{iar} \quad z_{0,06} = 1,56$$

În acest caz $z < z_{0,06}$ ipoteza nulă se acceptă.

Scanned with OKEN Scanner

11.2.2. Compararea mediilor a două eşantioane

În activitatea statistică se întâlnesc multe probleme în care trebuie să se decidă dacă diferenţa între mediile a două eşantioane poate fi atribuită întâmplării sau dacă există o indicaţie asupra faptului că eşantioanele au fost extrase din colectivităţi cu medii diferite. Să presupunem că ar trebui să se decidă dacă există diferenţe semnificative între siguranţa în funcţionare a două maşini de acelaşi tip, care produc acelaşi fel de produse sau aceeaşi operaţie, dacă o anumită operaţiune sau un anumit produs poate fi făcut mai bine de către o femeie sau de către un bărbat; dacă există diferenţe semnificative între nivelul de pregătire profesională a elevilor din două şcoli generale, una din mediul rural şi alta din mediul urban etc.

Metoda care se foloseşte pentru a testa dacă o diferenţă între două medii de sondaj se poate atribui întâmplării se bazează pe următorul principiu: dacă $\bar{x}_1$ şi $\bar{x}_2$ sunt mediile a două eşantioane independente de volum mare, n_1 şi n_2, distribuţia de sondaj a diferenţelor $\bar{x}_1 - \bar{x}_2$ poate fi considerată normală, cu media $m_1 - m_2$ şi abaterea standard:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (11.4)$$

unde m_1 şi m_2 reprezintă mediile celor două colectivităţi, din care s-au extras cele două eşantioane, şi σ_1^2 şi σ_2^2 dispersiile lor. Formula (11.4) reprezintă eroarea standard a diferenţei dintre două medii.

De multe ori în practică nu se cunosc dispersiile σ_1^2 şi σ_2^2. În aceste cazuri, dacă volumul eşantioanelor este suficient de mare ($n > 30$), este raţional să se înlocuiască dispersiile necunoscute cu dispersiile de sondaj s_1^2 şi s_2^2 şi, în aceste condiţii, testul ipotezei nule este dat de formula:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \qquad (11.5)$$

Scanned with OKEN Scanner

Din formula (11.5) rezultă că $z > 0$ când $\bar{x}_1 > \bar{x}_2$ şi $z < 0$, când $\bar{x}_1 < \bar{x}_2$. În funcţie de ipoteza alternativă $m_1 > m_2$, $m_1 < m_2$ sau $m_1 \neq m_2$, principiul pe care se bazează acest criteriu de semnificaţie se prezintă sugestiv în figura 11.1.

În tabelul 7.7 se prezintă două distribuţii de sondaj obţinute de la două ringuri 1 şi 2 cu privire la fineţea firelor de bumbac 100%. Pentru a verifica dacă cele două maşini funcţionează normal, s-au extras două sondaje, câte unul de la fiecare maşină, de volum $n_1 = n_2 = 240$ probe. Pe baza sondajelor s-au determinat $\bar{x}_1 = 51{,}266$; $\bar{x}_2 = 50{,}66$; $s_1^2 = 1{,}2989$ şi $s_2^2 = 1{,}1556$.

Pentru a verifica ipoteza diferenţei nule, adică a verifica dacă cele două maşini produc fire de aceeaşi fineţe, se va aplica criteriul (11.5). Considerând un nivel de semnificaţie $q_1 = 0{,}01$, rezultă:

$$0{,}01 = \frac{1}{2} - \Phi(z)$$

$$\Phi(z) = 0{,}49 \quad \text{deci} \quad z - 0{,}01 = -2{,}33$$

Substituind datele din problemă în relaţia (11.5) se obţine:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{51{,}266 - 50{,}66}{\sqrt{\frac{1{,}2989}{240} + \frac{1{,}1556}{240}}} = \frac{0{,}606}{0{,}101} = 6$$

Deoarece $z > z_{0{,}01}$, ipoteza diferenţei nule se acceptă; cu alte cuvinte, se poate considera că cele două eşantioane, extrase din producţia a două ringuri, provin din aceeaşi colectivitate.

În particular, dacă $\sigma_1^2 = \sigma_2^2$ sau respectiv $s_1^2 = s_2^2$, atunci problema se reduce la a compara între ele mediile a două eşantioane de volum n_1 şi n_2, având abaterile standard s_1 şi s_2 şi cu aceeaşi medie m. Problema care se pune în acest caz este dacă diferenţa dintre $\bar{x}_1$ şi $\bar{x}_2$ se datorează fluctuaţiilor de sondaj întâmplătoare sau unor diferenţe între sondaje. Menţionăm că cele două eşantioane trebuie să fie independente. În condiţiile sondajului simplu întâmplător, formula erorii standard a diferenţelor dintre medii devine:

Scanned with OKEN Scanner

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} \qquad (11.6)$$

Dacă nu se cunoaşte dispersia colectivităţii generale σ^2, aceasta se poate înlocui cu dispersiile de selecţie provenite din cele două eşantioane, astfel:

$$s_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}} \qquad (11.7)$$

Drept criteriu de verificare a diferenţei dintre mediile a două eşantioane extrase din aceeaşi colectivitate, se utilizează formula (11.7).

Să exemplificăm aplicarea acestui criteriu pe baza a două eşantioane de volum $n_1 = n_2 = 100$ extrase din cadrul colectivităţii celor 49808 mame distribuite pe grupe de vârstă din tabelul 10.2. Considerăm sondajul 37 şi sondajul 48. Pentru aceste sondaje se cunosc următoarele date:

Sondajul 37: $\begin{cases} \bar{x}_1 = 32{,}25 \\ s_1^2 = 26{,}69 \\ n_1 = 100 \end{cases}$ *Sondajul 48:* $\begin{cases} \bar{x}_2 = 30{,}85 \\ s_2^2 = 23{,}53 \\ n_2 = 100 \end{cases}$

La un nivel de semnificaţie $q_1 = 0{,}01$ rezultă $\Phi(z) = 0{,}49$ de unde $z_{-0{,}01} = -2{,}33$.

Substituind datele din problemă în formula (11.7) rezultă:

$$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{32{,}25-30{,}85}{\sqrt{\frac{26{,}69}{100}+\frac{23{,}53}{100}}} = 0{,}625$$

Deci $z > z_{-0{,}01}$ ceea ce înseamnă că ipoteza nulă se acceptă, adică diferenţa dintre cele două medii $\bar{x}_1 - \bar{x}_2 = 1{,}4$ ani este datorată fluctuaţiilor de sondaj.

Scanned with OKEN Scanner

11.2.3. Testarea erorii standard a medianei, cuartilelor şi decilelor

Am văzut, într-un capitol anterior al acestei cărţi, că atunci când seria pentru care se calculează indicatorii tendinţei centrale prezintă unele valori extreme ce se abat în mare măsură de la termenii din apropiere, mediana are o valoare mai stabilă şi mai puţin afectată de fluctuaţiile de selecţie decât media aritmetică.

Întrucât, de cele mai multe ori, în practica statistică mediana se calculează pe baza datelor provenite dintr-un sondaj, este necesar să cunoaştem un criteriu obiectiv cu ajutorul căruia să se poată aprecia în ce măsură mediana de sondaj poate fi considerată ca o estimaţie corectă a medianei colectivităţii generale din care provine sondajul.

Eroarea standard a medianei se determină cu ajutorul formulei:

$$s_{Me} = \sigma_x \sqrt{\frac{\pi}{2n}} \cong 1{,}25331\sigma_{\bar{x}} \qquad (11.8)$$

Vom exemplifica calculul şi analiza statistică a formulei (11.8), considerând sondajul 1 din cadrul colectivităţii mamelor căsătorite cu bărbaţi de vârstă 35-39 ani în 1975 (tabelul 11.3).

Vârsta mediană a mamelor din cadrul colectivităţii generale se va determina cu ajutorul formulei:

$$Me = x_0 + d\frac{\frac{\sum_{i=1}^{n} n_i}{2} - N_{ia}}{n_{Me}} = 30 - 5\frac{\frac{49808}{2} - 17830}{18629} = 31{,}9 \ ani$$

Deoarece distribuţia pe grupe de vârste a mamelor din colectivitatea generală se apropie de o distribuţie normală şi valoarea medianei se apropie de valoarea mediei aritmetice.

Vom calcula vârsta mediană pe baza sondajului 1, astfel:

$$Me = 30 - 5\frac{\frac{100}{2} - 36}{44} = 31{,}59 \ ani$$

Eroarea standard a medianei de sondaj calculată după formula (11.8), va fi:

Scanned with OKEN Scanner

$$s_{Me} = 1{,}25331 \cdot 0{,}491 = 0{,}62 \ ani$$

Aceasta înseamnă că, în două cazuri din trei, mediana colectivităţii generale se va încadra în intervalul:

$$(31{,}59 - 0{,}62) < Me < (31{,}59 + 0{,}62)$$

adică

$$30{,}97 < Me < 32{,}21$$

Altfel spus, vârsta mediană a celor 49808 mame se va încadra între limitele 30,97 ani şi 32,21 ani în 68 cazuri din 100 şi, evident, se va situa în afara acestor limite în 32 cazuri din 100.

Eroarea maximă admisibilă Δ_{Me} în cazul medianei de sondaj se va calcula pe baza formulei:

$$\Delta_{Me} = z \cdot s_{Me} \qquad (11.9)$$

în care z este argumentul funcţiei normale normate.

Astfel, când $z = 1$ $\Delta_{Me} = s_{Me}$ iar $P = 2\Phi(1) = 0{,}68268$

$z = 1{,}5$ $\Delta_{Me} = 1{,}5\, s_{Me}$ $P = 2\Phi(1{,}5) = 0{,}86638$

$z = 2{,}0$ $\Delta_{Me} = 2{,}0\, s_{Me}$ $P = 2\Phi(2{,}0) = 0{,}95450$

$z = 3{,}0$ $\Delta_{Me} = 3{,}0\, s_{Me}$ $P = 2\Phi(3{,}0) = 0{,}99730$

Dacă vom garanta rezultatele cu o probabilitate de 99,73 %, atunci intervalul în care se va cuprinde mediana colectivităţii generale va fi mai mare şi anume: $31{,}59 \pm 1{,}473$.

Eroarea standard a cuartilelor este aceeaşi pentru ambele cuartile şi se determină cu ajutorul formulei:

$$s_{Q_1} = s_{Q_2} = 1{,}36263 \cdot \sigma_{\bar{x}} \qquad (11.10)$$

iar *eroarea maximă a cuartilelor*, obţinute dintr-un sondaj, este dată de relaţia:

Scanned with OKEN Scanner

$$\Delta_{Q_1} = \Delta_{Q_2} = z \cdot s_{Q_1} = z \cdot s_{Q_2} \tag{11.11}$$

Erorile standard ale decilelor sunt egale pentru decilele echidistante şi se pot calcula cu ajutorul formulelor:

- pentru D_1 şi D_9: $s_{D_1} = s_{D_9} = 1{,}70942\ \sigma_{\bar{x}}$ (11.12)
- pentru D_2 şi D_8: $s_{D_2} = s_{D_8} = 1{,}42877\ \sigma_{\bar{x}}$ (11.13)
- pentru D_3 şi D_7: $s_{D_3} = s_{D_7} = 1{,}31800\ \sigma_{\bar{x}}$ (11.14)
- pentru D_4 şi D_6: $s_{D_4} = s_{D_6} = 1{,}26804\ \sigma_{\bar{x}}$ (11.15)

Din formulele erorilor standard ale medianei, cuartilelor şi decilelor, rezultă că influenţa fluctuaţiilor de selecţie asupra estimaţiilor acestor indicatori creşte pe măsură ce ne depărtăm de mediană, astfel încât eroarea standard a cuartilelor este mai mare decât cea a medianei cu aproape o zecime din $\sigma_{\bar{x}}$, iar eroarea standard a decilelor 1 şi 9 este mai mare cu aproape o treime din $\sigma_{\bar{x}}$.

11.2.4. Eroarea standard a abaterii medii pătratice şi a coeficientului de variaţie

Eroarea standard a abaterii medii pătratice de sondaj, în cazul unei distribuţii mesokurtice, se determină cu ajutorul formulei:

$$s_\sigma = \frac{\sigma_x}{\sqrt{2(n-1)}} \cong 0{,}707107\ \sigma_{\bar{x}} \tag{11.16}$$

Aplicând formula (11.16) în cazul sondajului 1 se obţine:

$$s_\sigma = \frac{\sigma_x}{\sqrt{2(n-1)}} = \frac{4{,}96}{\sqrt{2(100-1)}} = \frac{4{,}96}{14{,}07} = 0{,}35$$

Aceasta înseamnă că abaterea standard a colectivităţii generale va diferi de abaterea standard obţinută pe baza sondajului 1 cu ± 0,35 ani în două cazuri din trei.

Scanned with OKEN Scanner

Pentru o distribuţie oarecare, normală sau nenormală, eroarea standard se poate determina pe baza momentelor de diferite ordine cu ajutorul formulei:

$$s_\sigma = \sqrt{\frac{m_4 - m_2^2}{4m_2 \cdot n}} \qquad (11.17)$$

Vom exemplifica calculul şi analiza acestui indicator pe baza distribuţiei din sondajul 15. Calculele intermediare se prezintă în tabelul 11.3.

$$m_1 = \frac{\sum_{i=1}^{k}(x_i - x_0)n_i}{\sum_{i=1}^{k} n_i} = \frac{-60}{100} = -0{,}6$$

$$m_2 = \frac{\sum_{i=1}^{k}(x_i - x_0)^2 n_i}{\sum_{i=1}^{k} n_i} = \frac{2450}{100} = 24{,}5$$

$$m_4 = \frac{\sum_{i=1}^{k}(x_i - x_0)^4 n_i}{\sum_{i=1}^{k} n_i} = \frac{190250}{100} = 1902{,}5$$

Tabelul 11.3. Distribuţia mamelor pe grupe de vârstă şi calculul momentelor

Grupa de vârstă	n_i	Mijloacele intervalelor x_i	$d_i = x_i - x_0$	$d_i n_i$	$d_i^2 n_i$	$d_i^4 n_i$
15-19	1	17,5	-15	-15	225	50625
20-24	7	22,5	-10	-70	700	49000
25-29	24	27,5	-5	-120	600	15000
30-34	42	32,5	0	0	0	0
35-39	24	37,5	5	120	600	15000
40-44	1	42,5	10	10	100	10000
45-49	1	47,5	15	15	225	50625
Total	100	-	-	-60	2450	190250

Scanned with OKEN Scanner

Înlocuind aceste valori în (11.17) rezultă:

$$s_\sigma = \sqrt{\frac{m_4 - m_2^2}{4m_2 \cdot n}} = \sqrt{\frac{902{,}5 - 24{,}5^2}{4 \cdot 24{,}5 \cdot 100}} = 0{,}85 \text{ ani}$$

Rezultatele diferite care s-au obţinut pentru eroarea standard de sondaj prin aplicarea formulelor (11.16) şi (11.17), se datorează faptului că distribuţia de sondaj nu este normală.

Eroarea probabilă a unei abateri standard se calculează cu ajutorul formulei:

$$\Delta_s = z \cdot s_\sigma \qquad (11.18)$$

Eroarea standard a unui coeficient de variaţie se calculează cu ajutorul formulei:

$$s_\nu = \frac{\nu}{\sqrt{2n}} \sqrt{1 + 2\left(\frac{\nu}{100}\right)^2} \qquad (11.19)$$

Dacă coeficientul de variaţie este mai mic de 10%, atunci $\sqrt{1 + 2\left(\frac{\nu}{100}\right)^2}$ se apropie de unitate şi, prin urmare, se poate elimina din formula (11.19) fără a influenţa semnificativ rezultatul. Astfel, formula (11.19) devine:

$$s_\nu = \frac{\nu}{\sqrt{2n}} \qquad (11.20)$$

Eroarea probabilă sau eroarea maximă admisibilă a unui coeficient de variaţie devine:

$$\Delta_\nu = z \cdot s_\nu \qquad (11.21)$$

Intervalul de încredere pentru un coeficient de variaţie este:

$$\nu_s \pm \Delta_\nu$$

Scanned with OKEN Scanner

Utilizând aceeaşi distriibuţie pe grupe de vârstă a mamelor din tabelul 11.3 se obţine:

- în cazul colectivităţii generale: $v = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{4,96}{31,75} \cdot 100 = 15,62\%$

- în sondajul 10: $v = \frac{\sigma}{\bar{x}} \cdot 100 = \frac{4,83}{31,70} \cdot 100 = 15,24\%$

- eroarea standard a coeficientului de variaţie din sondaj va fi:

$$s_v = \frac{v}{\sqrt{2n}} \sqrt{1 + 2\left(\frac{v}{100}\right)^2} = \frac{0,1524}{\sqrt{2 \cdot 100}} \sqrt{1 + 2\left(\frac{0,1524}{100}\right)^2} = 0,0108$$

sau

$$s_v = \frac{v}{\sqrt{2n}} = \frac{0,1524}{\sqrt{2 \cdot 100}} = 0,0108 \quad \text{sau} \quad 1,08\%$$

Intervalul de încredere: $(v_s - \Delta_v) < v < (v_s + \Delta_v)$

La un nivel de semnificaţie de 1%, probabilitatea cu care se garantează rezultatele $P = 1 - 0,01 = 0,99$ iar din tabelele distribuţiei normale se găseşte $z = 2,58$. În aceste condiţii, eroarea limită va fi:

$$\Delta_v = \pm(2,58 \cdot s_v) = \pm(2,58 \cdot 1,08) = \pm 2,78\%$$

Deci $(15,24 - 2,78) < v < (15,24 + 2,78)$

Aceasta înseamnă că, în 99 de cazuri din 100, coeficientul de variaţie din colectivitatea generală se va situa între limitele 12,46 şi 18,02 şi numai într-un caz din 100 se va afla în afara acestor limite.

Scanned with OKEN Scanner

V. ANALIZA LEGĂTURILOR DINTRE FENOMENE

Scanned with OKEN Scanner

Scanned with OKEN Scanner

12. ANALIZA DISPERSIONALĂ

13. LEGĂTURILE DINTRE FENOMENELE REALE

14. REGRESIE ŞI CORELAŢIE SIMPLĂ LINIARĂ

15. REGRESIE ŞI CORELAŢIE SIMPLĂ CURBILINIE

16. REGRESIE ŞI CORELAŢIE MULTIPLĂ LINIARĂ

17. REGRESIE ŞI CORELAŢIE MULTIPLĂ CURBILINIE

18. CORELAŢIE NEPARAMETRICĂ

19. ASOCIAŢIA ÎNTRE CARACTERISTICI CALITATIVE

20. ELASTICITATEA

21. FUNCŢIILE DE PRODUCŢIE

12. ANALIZA DISPERSIONALĂ

Metoda analizei dispersionale, ca şi termenul *analiză dispersională*, cunoscută şi sub denumirea de *analiză a varianţei*, provine de la *ANalyse Of VAriance* prescurtat *ANOVA*, şi au fost introduse în ştiinţă de către statisticianul şi matematicianul englez *Ronald Fisher* [60] , colaborator şi succesor al lui *K. Pearson* la catedra de specialitate a Universităţii din Londra.

ANOVA reprezintă o metodă de studiere a influenţei factorilor care acţionează simultan asupra unui fenomen colectiv, pe baza unor date experimentale.

La *Staţiunea experimentală Rothamsted, R. A. Fisher* a fost confruntat cu probleme de felul următor: compararea producţiilor medii ale unor varietăţi de cereale şi plante tehnice, cultivate pe anumite suprafeţe de teren; determinarea influenţei unor tipuri de îngrăşăminte asupra volumului producţiei etc.

Din punct de vedere statistic, aceste tipuri de probleme se reduc la compararea mediilor unor colectivităţi diferite şi la verificarea ipotezei statistice asupra omogenităţii acestora.

La baza analizei dispersionale stă metoda grupării variantelor variabilelor statistice după criterii variate.

ANOVA constă în *descompunerea variaţiei totale* (a varianţei totale) în mai multe componente (cel puţin două), reflectând influenţele exercitate de către diferiţi factori asupra fenomenului studiat. *Fisher* a estimat dispersia unei caracteristici supuse influenţei unui factor, iar apoi, înlăturând influenţa acestuia şi comparând cele două dispersii, a obţinut informaţii referitoare la influenţa factorului respectiv asupra caracteristicii cercetate. Factorul respectiv exercită o influenţă importantă asupra (caracteristicii) fenomenului studiat, dacă prima dispersie este mai mare decât cea de a doua şi invers.

ANOVA poate fi considerată o generalizare a testului t pentru compararea a mai mult de două medii ale unor caracteristici normal distribuite.

În funcţie de numărul de factori, ale căror influenţă se studiază, deosebim:

- *analiză dispersională unifactorială*;
- *analiză dispersională multifactorială*.

Atât analiza dispersională unifactorială, cât şi cea multifactorială se pot realiza fie pe baza unor eşantioane de acelaşi volum n,fie pe baza unor eşantioane de volum diferit.

Scanned with OKEN Scanner

12.1. ANOVA unifactorială

12.1.1. ANOVA unifactorială cu eşantioane de volum n

Fie k un număr de subcolectivităţi, care au ipotetic aceeaşi medie m a caracteristicii X. Din fiecare subcolectivitate se extrage câte un eşantion aleator de acelaşi volum n, cu scopul de a vedea dacă mediile acestora $\bar{x}_j$ diferă semnificativ de m.

Datele iniţiale se pot sistematiza într-un tabel cu dublă intrare (tabelul 12.1), în care pe coloane se înscriu cele k eşantioane (tratamente), iar pe linii variantele caracteristicii X din fiecare sondaj i. Astfel, x_{ij} va reprezenta varianta i din sondajul j al caracteristicii X.

Tabelul 12.1. Sistematizarea datelor empirice în cazul ANOVA unifactorială

Tratamentul / Variante ale caracteristicii	1	2	3	...	j	...	k	Total
1	x_{11}	x_{12}	x_{13}	...	x_{1j}	...	x_{1k}	$\sum_{j=1}^{k} x_{1j}$
2	x_{21}	x_{22}	x_{23}	...	x_{2j}	...	x_{2k}	$\sum_{j=1}^{k} x_{2j}$
3	x_{31}	x_{32}	x_{33}	...	x_{3j}	...	x_{3k}	$\sum_{j=1}^{k} x_{3j}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
i	x_{i1}	x_{i2}	x_{i3}	...	x_{ij}	...	x_{ik}	$\sum_{j=1}^{k} x_{ij}$
⋮	⋮	⋮	⋮		⋮		⋮	⋮
n	x_{n1}	x_{n2}	x_{n3}		x_{nj}		x_{nk}	$\sum_{j=1}^{n} x_{nj}$
Total	$\sum_{i=1}^{n} x_{i1}$	$\sum_{i=1}^{n} x_{i2}$	$\sum_{i=1}^{n} x_{i3}$	...	$\sum_{i=1}^{n} x_{ij}$	...	$\sum_{i=1}^{n} x_{ik}$	$\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}$
Mediile	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	...	$\bar{x}_j$	...	$\bar{x}_k$	$\bar{x}$

Scanned with OKEN Scanner

ANOVA presupune descompunerea variaţiei totale în cel puţin două componente şi apoi estimarea dispersiilor corespunzătoare şi compararea acestora.

În majoritatea aplicaţiilor practice, cele k sondaje aleatoare se consideră că reflectă influenţa unui anumit factor (tratament) care acţionează asupra fenomenului colectiv studiat sub aspectul caracteristicii X. În acest fel, se separă influenţa factorilor esenţiali de influenţele celorlalţi factori, consideraţi întâmplători.

În acest tabel s-a notat:

- totalul caracteristicii X din fiecare sondaj cu $\sum_{i=1}^{n} x_{ij}$ $\quad i=\overline{1,n}$ şi $j=\overline{1,k}$

- totalul general $\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}$

- mediile sondajelor $\bar{x}_j = \dfrac{\sum_{i=1}^{n} x_{ij}}{n}$ (12.1)

- media generală $\bar{x} = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}}{nk}$ (12.2)

Variaţia fiecărei variante a caracteristicii x_{ij} faţă de media generală $\bar{x}$ se poate descompune în două componente, astfel:

$$x_{ij} - \bar{x} = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{x}) \tag{12.3}$$

Ridicând la pătrat relaţia (12.3) şi însumând rezultă:

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2 =$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2 + 2\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}) + \sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2 \tag{12.4}$$

unde

$$2\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})=2\sum_{j=1}^{k}\left[\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)(\bar{x}_j-\bar{x})\right]=$$

$$2\sum_{j=1}^{k}\left[(\bar{x}_j-\bar{x})\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)\right]=0$$

deci

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2=\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2+\sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2$$

şi

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2=\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2+n\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2 \qquad (12.5)$$

Relaţia (12.5) arată că *variaţia totală* sau *varianţa totală* (notată cu SPA_t) a caracteristicii X din cele k eşantioane de volum n a fost descompusă în două componente principale:

- *variaţia dintre eşantioane* (dintre grupe), numită şi variaţie sau *varianţă sistematică* sau *factorială* (datorată factorului de grupare) notată cu SPA_f:

$$SPA_f=n\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2 \qquad (12.6)$$

- *variaţia (varianţa) reziduală* sau *rămasă*, calculată ca o variaţie din interiorul (grupelor) eşantioanelor, notată cu SPA_r:

$$SPA_r=\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2 \qquad (12.7)$$

rezultă

$$SPA_t=SPA_f+SPA_r \qquad (12.8)$$

Al doilea termen din membrul doi al relaţiei (12.8), suma pătratelor abaterilor variantelor caracteristicii din fiecare sondaj faţă de media lui, devine:

Scanned with OKEN Scanner

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x}_j)^2=\sum_{i=1}^{n}(x_{i1}-\bar{x}_1)^2+\sum_{i=1}^{n}(x_{i2}-\bar{x}_2)^2+\cdots+\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2$$

Această abatere este o măsură a variaţiei în interiorul sondajelor faţă de mediile lor respective, este independentă de orice diferenţă dintre sondaje şi depinde numai de factorii întâmplători, care au acţionat în cazul fiecărui sondaj, am putea spune în cadrul fiecărui experiment şi de aceea se mai numeşte *abatere experimentală* sau *eroare experimentală* sau *abatere reziduală.*

Primul termen din membrul al doilea al relaţiei (12.5) reprezintă suma pătratelor abaterilor mediilor de sondaj $\bar{x}_j$ faţă de media generală $\bar{x}$ şi măsoară varianţa datorată factorului de sondaj (sau a tratamentelor). Astfel, cu cât diferenţele dintre mediile de sondaj sunt mai mici, cu atât suma pătratelor abaterilor respective va fi mai mică şi invers. Această variaţie poartă denumirea de *variaţie factorială* (*esenţială*).

Calculul celor trei sume de pătrate ale abaterilor din relaţia (12.5) se poate face mai uşor în practică pe baza unor formule transformate. Astfel:

- *varianţa totală:*

$$SPA_t=\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij}-\bar{x})^2=\sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}^2-\frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}\right)^2}{kn} \qquad (12.9)$$

- *varianţa factorială:*

$$SPA_f=n\sum_{j=1}^{k}(\bar{x}_j-\bar{x})^2=n\left[\sum_{j=1}^{k}\bar{x}_j^2-\frac{\left(\sum_{j=1}^{k}\bar{x}_j\right)^2}{k}\right]=$$

$$= n\left[\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{x_{ij}^2}{n^2} - \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}\right)^2}{kn^2}\right] = \sum_{i=1}^{n}\sum_{j=1}^{k}\frac{x_{ij}^2}{n} - \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}\right)^2}{kn} \qquad (12.10)$$

• *varianţa reziduală:*

$$SPA_r = \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2 = \sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}^2 - \frac{\left(\sum_{i=1}^{n}\sum_{j=1}^{k}x_{ij}\right)^2}{n} \qquad (12.11)$$

Considerând că cele k eşantioane ar fi extrase din aceeaşi colectivitate, dispersia acesteia se poate estima în trei feluri:

1. Ca *dispersie reziduală*, corectată, calculată prin împărţirea SPA_r la numărul gradelor de libertate $k(n-1)$:

$$s_r^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2}{k(n-1)} \qquad (12.12)$$

2. Ca *dispersie factorială*, corectată, calculată prin împărţirea SPA_f la numărul gradelor de libertate corespunzător $k-1$:

$$s_f^2 = \frac{n\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2}{k-1} \qquad (12.13)$$

3. *Dispersia totală* se determină prin împărţirea SPA_t la numărul gradelor de libertate $nk-1$:

$$s_t^2 = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2}{nk-1} \qquad (12.14)$$

Între gradele de libertate utilizate în cele trei dispersii se verifică relaţia:

$$nk - 1 = (k-1) + k(n-1) \qquad (12.15)$$

Analiza dispersională constă în compararea celor două estimaţii ale dispersiei generale s_f^2 şi s_r^2. Dacă cele două estimaţii nu diferă semnificativ una de cealaltă, atunci raportul dintre ele, coeficientul lui *Fisher* F_c, este apropiat de 1. Dacă toate seturile posibile de k eşantioane de volum n sunt extrase dintr-o colectivitate normal distribuită, atunci raportul dintre cele două estimaţii ale dispersiei generale urmează o distribuţie F cu $\nu_1 = k-1$ şi $\nu_2 = k(n-1)$ grade de libertate.

Calculele intermediare necesare pentru aplicarea metodei analizei dispersionale unifactoriale se pot sistematiza sub forma tabelului 12.2.

Tabelul 12.2. *ANOVA* unifactorială în cazul unor eşantioane de acelaşi volum n

Natura variaţiei	Suma pătratelor abaterilor	Numărul gradelor de libertate	Dispersia	$F_{calculat}$	$F_{tabelar}$
Variaţia factorială	$n\sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2$	$\nu_1 = k-1$	s_f^2	$F_c = \frac{s_f^2}{s_r^2}$	F_{α,ν_1,ν_2}
Variaţia reziduală	$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2$	$\nu_2 = k(n-1)$	s_r^2		
Variaţia totală	$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2$	$\nu = nk-1$	s_t^2		

Scanned with OKEN Scanner

Dacă cele k eşantioane sunt extrase din colectivităţi cu medii diferite atunci dispersia factorială va fi considerabil mai mare decât dispersia reziduală iar valoarea calculată $F_{calculat}$ va fi mai mare decât $F_{tabelar}$ pentru un nivel de semnificaţie $\rho = 1 - P$.

Algoritmul de calcul care se aplică în practică în mod obişnuit este:

1) Se determină :

$$F_c = \frac{s_f^2}{s_r^2} \tag{12.16}$$

2) În funcţie de nivelul de semnificaţie ρ ales şi de numerele gradelor de libertate aferente varianţei factoriale ν_1 şi celei reziduale ν_2, se caută valoarea (teoretică) tabelară a coeficientului lui *Fisher* F_{ρ,ν_1,ν_2} în tabelele cu valorile critice ale repartiţiei F.

Comparând valoarea lui F, obţinută prin calcul pe baza formulei (12.16), cu valoarea tabelară F_t corespunzătoare, rezultă:

$$F_c < F_t \quad \text{sau} \quad F_c = F_t \quad \text{sau} \quad F_c > F_t \tag{12.17}$$

Rezultatul acestei comparaţii (12.17) se interpretează astfel:

- dacă $F_c < F_t$, se acceptă ipoteza omogenităţii mediilor eşantioanelor, ceea ce înseamnă că tratamentele efectuate (factorul de grupare) nu au avut influenţă asupra mediilor eşantioanelor.
- dacă $F_c \geq F_t$, se respinge ipoteza omogenităţii mediilor eşantioanelor, deci se acceptă ipoteza eterogenităţii mediilor şi rezultă că factorul de grupare a avut o influenţă importantă asupra mediilor eşantioanelor.

Exemplu. Pentru a caracteriza modul de funcţionare a cinci maşini de acelaşi tip, care produc acelaşi produs, s-au extras cinci eşantioane de volum egal cu cinci probe de pe fiecare maşină (tabelul 12.3).

Scanned with OKEN Scanner

Tabelul 12.3

Eşantionul (j) / Varianta (i)	1 2 3 4 5	Total
1 2 3 4 5	49 52 49 51 50 51 53 52 48 49 50 51 49 49 53 51 55 52 50 47 54 54 53 52 51	
Total $\sum_{i=1}^{n} x_{ij}$	255 265 255 250 250	$1275 = \sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}$
$\bar{x}_j$ $\bar{x}_j - \bar{x}$ $(\bar{x}_j - \bar{x})^2$	51 53 51 50 50 0 2 0 -1 -1 0 4 0 1 1	$51 = \bar{x}$ $5 \cdot 6 = \sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2$ 30
$x_{1j} - \bar{x}_j$ $x_{2j} - \bar{x}_j$ $x_{3j} - \bar{x}_j$ $x_{4j} - \bar{x}_j$ $x_{5j} - \bar{x}_j$	-2 -1 -2 1 0 0 0 1 -2 -1 -1 -2 -2 -1 3 0 2 1 0 -3 3 1 2 2 1	
$(x_{1j} - \bar{x}_j)^2$ $(x_{2j} - \bar{x}_j)^2$ $(x_{3j} - \bar{x}_j)^2$ $(x_{4j} - \bar{x}_j)^2$ $(x_{5j} - \bar{x}_j)^2$	4 1 4 1 0 0 0 1 4 1 1 4 4 1 9 0 4 1 0 9 9 1 4 4 1	10 6 19 14 19
$\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$	14 10 14 10 20	$68 = \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2$
$(x_{1j} - \bar{x})^2$ $(x_{2j} - \bar{x})^2$ $(x_{3j} - \bar{x})^2$ $(x_{4j} - \bar{x})^2$ $(x_{5j} - \bar{x})^2$	4 1 4 0 1 0 4 1 9 4 1 0 4 4 4 0 16 1 1 16 9 9 4 1 0	10 18 13 34 23
$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2$	14 30 14 15 25	$98 = \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2$

Scanned with OKEN Scanner

Tabelul 12.4. Schema *ANOVA* unifactoriale cu eşantioane egale

Natura variaţiei	Suma pătratelor abaterilor	Numărul gradelor de libertate	Dispersia	$F_{calculat}$	$F_{tabelar}$
Variaţia factorială	30	$\nu_1 = 5-1$	$s_f^2 = \frac{30}{4} = 7,5$	$\frac{s_f^2}{s_r^2} = \frac{7,5}{3,2}$	$F_{\alpha,\nu_1,\nu_2} = 2,87$
Variaţia reziduală	68	$\nu_2 = 5 \cdot 4$	$s_r^2 = \frac{68}{20} = 3,2$	$F_c = 2,21$	
Variaţia totală	98	$\nu = 55-1$	$s^2 = \frac{98}{24} = 4,5$	$F_c < F_t$ se admite ipoteza de omogenitate	

Deoarece $F_c < F_t$, rezultă că ipoteza privind omogenitatea celor cinci eşantioane se confirmă, cu un nivel de semnificaţie de 5%, deci nu există o diferenţă semnificativă între regimurile de funcţionare ale celor 5 maşini.

12.1.2. Analiza dispersională unifactorială pe baza unor eşantioane de volum diferit

Considerăm p colectivităţi de volum N_j, din care extragem p eşantioane independente de volum $n_1, n_2, \cdots, n_i, \cdots, n_p$ astfel:

$$\begin{array}{ccccccc} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{array}$$

$$n_1 \neq n_2 \neq \cdots \neq n_j \neq \cdots \neq n_p$$

Volumul tuturor eşantioanelor : $n = \sum_{j=1}^{p} n_j \qquad j = \overline{1, p}$

Media sondajului j : $\bar{x}_j = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}}{n_j}$

Media generală : $\bar{x} = \dfrac{\sum_{i=1}^{n}\sum_{j=1}^{k} x_{ij}}{n}$

Variaţia variantelor x_{ij} faţă de media generală se poate scrie:

$$x_{ij} - \bar{x} = (\bar{x}_j - \bar{x}) + (x_{ij} - \bar{x}_j) \tag{12.18}$$

Variaţia totală s-a descompus în două componente principale:

- *variaţia factorială* : $(\bar{x}_j - \bar{x})$ şi
- *variaţia reziduală* : $(x_{ij} - \bar{x}_j)$

Ridicând la pătrat şi însumând elementele relaţiei (12.18), se obţine ecuaţia analizei dispersionale:

$$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2 = \sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2 n_j + 2\sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{x}_j - \bar{x})(x_{ij} - \bar{x}_j) + \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2$$

dar $$2\sum_{i=1}^{n}\sum_{j=1}^{k}(\bar{x}_j - \bar{x})(x_{ij} - \bar{x}_j) = 0$$

rezultă $$\sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x})^2 = \sum_{j=1}^{k}(\bar{x}_j - \bar{x})^2 n_j + \sum_{i=1}^{n}\sum_{j=1}^{k}(x_{ij} - \bar{x}_j)^2$$

Scanned with OKEN Scanner

Deci, suma pătratelor abaterilor totale SPA_t este împărţită în două componente aditive: *suma pătratelor abaterilor factoriale* sau între eşantioane SPA_f şi *suma pătratelor abaterilor reziduale* sau din interiorul eşantioanelor SPA_r.

Relaţia (12.19) devine $$SPA_t = SPA_f + SPA_r \tag{12.20}$$

Pentru verificarea ipotezelor considerăm că cele p eşantioane s-au extras din p colectivităţi normal distribuite, având aceeaşi abatere standard σ, iar eşantioanele sunt aleatoare simple şi independente unele de altele.

Împărţind sumele de pătrate ale abaterilor din relaţia (12.20) prin numerele gradelor de libertate corespunzătoare, se obţin trei varianţe (pătrate medii PM) sau trei dispersii astfel:

$$PM_t = \frac{SPA_t}{n-1} \tag{12.21}$$

$$PM_f = \frac{SPA_f}{p-1} \tag{12.22}$$

$$PM_r = \frac{SPA_r}{n-p} \tag{12.23}$$

Între gradele de libertate corespunzătoare celor trei varianţe există următoarea relaţie:

$$n-1 = (p-1) + (n-p) \tag{12.24}$$

Când ipoteza nulă este adevărată, rezultă:

$$\chi_t^2 = \frac{SPA_t}{\sigma^2}; \quad \chi_r^2 = \frac{SPA_r}{\sigma^2}; \quad \chi_f^2 = \frac{SPA_f}{\sigma^2}$$

şi $$\frac{SPA_t}{\sigma^2} = \frac{SPA_r}{\sigma^2} + \frac{SPA_f}{\sigma^2}$$

deci $$\chi_t^2 = \chi_r^2 + \chi_f^2 \tag{12.25}$$

Scanned with OKEN Scanner

iar $$F_{calc} = \frac{\chi_f^2/(p-1)}{\chi_r^2/(n-p)} = \frac{SPA_f/(p-1)}{SPA_r/(n-p)} = \frac{PM_f}{PM_r} \qquad (12.26)$$

Tabelul 12.5. Schema analizei dispersionale unifactoriale cu eşantioane de volum diferit

Sursa variaţiei	Grade de libertate	Suma pătratelor abaterilor	Dispersiile	F
1	2	3	4	5
Factorială	$p-1$	SPA_f	PM_f	$F_c = \frac{PM_f}{PM_r}$
Reziduală	$n-p$	SPA_r	PM_r	F_{ρ,ν_1,ν_2}
Totală	$n-1$	SPA_t	PM_t	

Când ipoteza nulă se acceptă (se verifică): $F_c < F_t$, F_c este deci o valoare observată a unei variabile F a lui *Snedecor*, cu $k_1 = p-1$ şi $k_2 = n-p$ grade de libertate. Media acestei variabile este egală cu unitatea.

În cazul opus, când *ipoteza nulă se respinge* mărimea variabilei F este superioară unităţii şi anume, cu atât mai mare, cu cât ipoteza nulă trebuie mai mult respinsă. Corespunzător unui anumit nivel de semnificaţie, ipoteza nulă se respinge atunci când:

$$F_{calc} \geq F_{\rho,k_1,k_2}$$

Dacă $F_{calc} < F_{\rho,k_1,k_2}$, se admite ipoteza nulă adică omogenitatea mediilor eşantioanelor colectivităţii, asumându-se un risc $\rho = 1-P$, unde P este probabilitatea.

Scanned with OKEN Scanner

12.2. ANOVA bifactorială

12.2.1. ANOVA bifactorială cu o singură observaţie în fiecare grup

Să presupunem că asupra unui fenomen X acţionează doi factori A şi B, iar observaţiile sunt grupate combinat după cei doi factori, având o singură observaţie în fiecare grupă, ceea ce înseamnă un total de $n = pq$ observaţii, conform tabelului 12.6.

Tabelul 12.6. Distribuţia fenomenului X după factorii A şi B

A \ B	1	2	...	l	...	q	$\bar{x}_{k.}$
1	x_{11}	x_{12}	...	x_{1l}	...	x_{1q}	$\bar{x}_{1.}$
2	x_{21}	x_{22}	...	x_{2l}	...	x_{2q}	$\bar{x}_{2.}$
3	x_{31}	x_{32}	...	x_{3l}	...	x_{3q}	$\bar{x}_{3.}$
⋮	.....						⋮
k	x_{k1}	x_{k2}	...	x_{kl}	...	x_{kq}	$\bar{x}_{k.}$
⋮	.....						⋮
p	x_{p1}	x_{p2}	...	x_{pl}	...	x_{pq}	$\bar{x}_{p.}$
$\bar{x}_{.}$	$x_{.1}$	$x_{.2}$	...	$x_{.l}$	...	$x_{.q}$	$\bar{x}_{..}$

Nivelul mediu al fenomenului X, condiţionat de fiecare din cei doi factori A şi B va fi:

$$\bar{x}_{k.} = \frac{\sum_{l=1}^{q} x_{kl}}{q} \tag{12.27}$$

şi

$$\bar{x}_{.l} = \frac{\sum_{k=1}^{p} x_{kl}}{p} \tag{12.28}$$

Scanned with OKEN Scanner

iar media generală $$\bar{x}_{..} = \frac{\sum_{k=1}^{p}\sum_{l=1}^{q} x_{kl}}{pq} \qquad (12.29)$$

Variaţia totală a variabilei X faţă de media generală se poate descompune în trei componente:

$$x_{kl} - \bar{x}_{..} = x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} + \bar{x}_{..} + \bar{x}_{k.} - \bar{x}_{..} + \bar{x}_{.l} - \bar{x}_{..} =$$

$$= (x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} + \bar{x}_{..}) + (\bar{x}_{k.} - \bar{x}_{..}) + (\bar{x}_{.l} - \bar{x}_{..})$$

Ridicând la pătrat şi însumând relaţia anterioară devine:

$$\sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{..})^2 = \sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} + \bar{x}_{..})^2 + q\sum_{k=1}^{p}(\bar{x}_{k.} - \bar{x}_{..})^2 + p\sum_{l=1}^{q}(\bar{x}_{.l} - \bar{x}_{..})^2 \qquad (12.30)$$

Tabelul 12.7. Schema analizei dispersionale bifactoriale cu o singură observaţie în fiecare grup

Natura împrăştierii	Suma pătratelor abaterilor	Numărul gradelor de libertate	Dispersia	$F_{calculat}$	$F_{tabelar}$
Variaţia determinată de factorul A	$q\sum_{k=1}^{p}(\bar{x}_{k.} - \bar{x}_{..})^2$	$f_1 = p-1$	s_1^2	$F_{c1} = \frac{s_1^2}{s_3^2}$	F_{p,f_1,f_3}
Variaţia determinată de factorul B	$p\sum_{l=1}^{q}(\bar{x}_{.l} - \bar{x}_{..})^2$	$f_2 = q-1$	s_2^2	$F_{c2} = \frac{s_2^2}{s_3^2}$	F_{p,f_2,f_3}
Variaţia reziduală	$\sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} - \bar{x}_{..})^2$	$f_3 = (p-1)(q-1)$	s_3^2		
Variaţia totală	$\sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{..})^2$	$pq-1$			

Scanned with OKEN Scanner

Împărţind aceste sume de pătrate (12.30) la numărul gradelor de libertate corespunzătoare, rezultă dispersiile:

- dintre linii $$s_1^2 = \frac{q\sum_{k=1}^{p}(\bar{x}_{k.} - \bar{x}_{..})^2}{p-1} \qquad (12.31)$$

- dintre coloane $$s_2^2 = \frac{p\sum_{l=1}^{q}(\bar{x}_{.l} - \bar{x})^2}{q-1} \qquad (12.32)$$

- reziduală $$s_3^2 = \frac{\sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{k.} - \bar{x}_{.l} + \bar{x}_{..})^2}{(p-1)(q-1)} \qquad (12.33)$$

- totală $$s^2 = \frac{\sum_{k=1}^{p}\sum_{l=1}^{q}(x_{kl} - \bar{x}_{..})^2}{pq-1} \qquad (12.34)$$

Între gradele de libertate se verifică relaţia:

$$pq - 1 = (p-1) + (q-1) + (p-1)(q-1) \qquad (12.35)$$

Dacă $F_{c_1} < F_{p,f_1,f_3}$, nu se admite existenţa unei influenţe reale a factorului A şi invers.

Dacă $F_{c_2} < F_{p,f_2,f_3}$, nu se admite existenţa unei influenţe importante a factorului B şi invers.

12.2.2. ANOVA bifactorială cu eşantioane de acelaşi volum

Să presupunem că din pq colectivităţi de fenomene X, supuse acţiunii simultane a doi factori A şi B se extrag în mod aleator pq sondaje de acelaşi volum n. Fiecare colectivitate are caracteristica X normal distribuită cu media m şi dispersia σ.

Scanned with OKEN Scanner

Tabelul 12.8. *ANOVA* bifactorială în cazul unor eşantioane de acelaşi volum

A	1		...				...	i						...	p						Total
K \ B	1	2	...	j	...	q	...	1	2	...	j	...	q	...	1	2	...	j	...	q	
1	x_{111}	x_{121}	...	x_{1j1}	...	x_{1q1}	...	x_{i11}	x_{i21}	...	x_{ij1}	...	x_{iq1}	...	x_{p11}	x_{p21}	...	x_{pj1}	...	x_{pq1}	
2	x_{112}	x_{122}	...	x_{1j2}	...	x_{1q2}	...	x_{i12}	x_{i22}	...	x_{ij2}	...	x_{iq2}	...	x_{p12}	x_{p22}	...	x_{pj2}	...	x_{pq2}	
⋮	⋮	⋮	...	⋮	...	⋮	...	⋮	⋮	...	⋮	...	⋮	...	⋮	⋮	...	⋮	...	⋮	
k	x_{11k}	x_{12k}	...	x_{1jk}	...	x_{1qk}	...	x_{i1k}	x_{i2k}	...	x_{ijk}	...	x_{iqk}	...	x_{p1k}	x_{p2k}	...	x_{pjk}	...	x_{pqk}	
⋮	⋮	⋮	...	⋮	...	⋮	...	⋮	⋮	...	⋮	...	⋮	...	⋮	⋮	...	⋮	...	⋮	
n	x_{11n}	x_{12n}	...	x_{1jn}	...	x_{1qn}	...	x_{i1n}	x_{i2n}	...	x_{ijn}	...	x_{iqn}	...	x_{p1n}	x_{p2n}	...	x_{pjn}	...	x_{pqn}	
$\sum_{k=1}^{n} x_{ijk}$	$x_{11.}$	$x_{12.}$	...	$x_{1j.}$	...	$x_{1q.}$	...	$x_{i1.}$	$x_{i2.}$	...	$x_{ij.}$	...	$x_{iq.}$	...	$x_{p1.}$	$x_{p2.}$	...	$x_{pj.}$	...	$x_{pq.}$	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n} x_{ijk}$
$\sum_{k=1}^{n} x_{ijk}^2$	$x_{11.}^2$	$x_{12.}^2$	...	$x_{1j.}^2$	...	$x_{1q.}^2$	...	$x_{i1.}^2$	$x_{i2.}^2$	...	$x_{ij.}^2$	...	$x_{iq.}^2$	...	$x_{p1.}^2$	$x_{p2.}^2$	...	$x_{pj.}^2$	...	$x_{pq.}^2$	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n} x_{ijk}^2$
$\bar{x}_{ij.}$	$\bar{x}_{11.}$	$\bar{x}_{12.}$	...	$\bar{x}_{1j.}$	...	$\bar{x}_{1q.}$	...	$\bar{x}_{i1.}$	$\bar{x}_{i2.}$	...	$\bar{x}_{ij.}$	...	$\bar{x}_{iq.}$	...	$\bar{x}_{p1.}$	$x_{p2.}$	...	$\bar{x}_{pj.}$	...	$\bar{x}_{pq.}$	

Scanned with OKEN Scanner

$$\bar{x}_{ij.} = \frac{\sum_{k=1}^{n} x_{ijk}}{n} \qquad i = \overline{1,p} \quad j = \overline{1,q} \tag{12.36}$$

Media variantelor caracteristicii după primul criteriu de clasificare:

$$\bar{x}_{i..} = \frac{1}{qn_j} \sum_{j=1}^{q} \sum_{k=1}^{n} x_{ijk} \qquad i = \overline{1,p} \tag{12.37}$$

sau

$$\bar{x}_{i..} = \frac{1}{q} \sum_{j=1}^{q} \bar{x}_{ij} \tag{12.38}$$

Media variantelor caracteristicii după al doilea criteriu de clasificare:

$$\bar{x}_{.j.} = \frac{1}{np} \sum_{i=1}^{p} \sum_{j=1}^{q} x_{ijk} \qquad j = \overline{1,q} \tag{12.39}$$

sau

$$\bar{x}_{.j.} = \frac{1}{p} \sum_{i=1}^{p} \bar{x}_{ij} \tag{12.40}$$

Media generală a tuturor eşantioanelor:

$$\bar{x}_{...} = \frac{1}{pq} \sum_{i=1}^{p} \sum_{j=1}^{q} x_{ij.} \qquad i = \overline{1,p} \quad j = \overline{1,q} \tag{12.41}$$

Abaterea variantelor caracteristicii faţă de media lor generală pentru toate sondajele se poate scrie:

$$x_{ijk} - \bar{x}_{...} = (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...}) + (\bar{x}_{i..} - \bar{x}_{...}) + (\bar{x}_{.j.} - \bar{x}_{...}) + (x_{ijk} - \bar{x}_{ij.}) \tag{12.42}$$

Ridicând la pătrat şi însumând după cei trei indici i, j, k, rezultă:

Scanned with OKEN Scanner

$$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{...})^2 = n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x}_{...})^2 + nq\sum_{i=1}^{p}(\bar{x}_{i..}-\bar{x}_{...})^2 +$$

$$+pn\sum_{j=1}^{p}(\bar{x}_{.j.}-\bar{x}_{...})^2 + \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij.})^2 \qquad (12.43)$$

În relaţia (12.43), membrul întâi reprezintă suma pătratelor abaterilor valorilor observate x_{ijk} faţă de media lor SPA_t, iar ultimul termen din membrul al doilea reprezintă suma pătratelor abaterilor SPA_c din interiorul fiecărei celule i, j.

Raportând cele două sume de pătrate la numărul gradelor de libertate corespunzător, rezultă *dispersiile*:

- *totală*: $$s^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{...})^2}{npq-1} \qquad (12.44)$$

- *reziduală* : $$s_4^2 = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij.})^2}{pq(n-1)} \qquad (12.45)$$

- *interacţiunea dintre factorii A şi B:*

$$s_3^2 = \frac{n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x}_{...})^2}{(p-1)(q-1)} \qquad (12.46)$$

- *influenţa factorului A* : $$s_1^2 = \frac{nq\sum_{i=1}^{p}(\bar{x}_{i..}-\bar{x}_{...})^2}{p-1} \qquad (12.47)$$

Scanned with OKEN Scanner

- *influenţa factorului B* : $$s_2^2 = \frac{pn\sum_{j=1}^{q}(\bar{x}_{.j.} - \bar{x}_{...})^2}{(q-1)} \qquad (12.48)$$

Între numitorii celor cinci dispersii se verifică relaţia:

$$pqn - 1 = pq(n-1) + (p-1)(q-1) + (p-1) + (q-1) \qquad (12.49)$$

Tabelul 12.9. Schema analizei dispersionale bifactoriale în cazul eşantioanelor cu mai multe observaţii de acelaşi volum

Natura variaţiei	Suma pătratelor	Numărul gradelor de libertate	Disper--siile	$F_{calculat}$	$F_{tabelar}$
Factorul A	$nq\sum_{i=1}^{p}(\bar{x}_{i..} - \bar{x}_{...})^2$	$f_1 = p-1$	s_1^2	$F_1 = \frac{s_1^2}{s_4^2}$	F_{p,f_1,f_4}
Factorul B	$np\sum_{j=1}^{q}(\bar{x}_{.j.} - \bar{x}_{...})^2$	$f_2 = q-1$	s_2^2	$F_2 = \frac{s_2^2}{s_4^2}$	F_{p,f_2,f_4}
Interacţiu-nea AB	$n\sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{x}_{...})^2$	$f_3 = (q-1)(p-1)$	s_3^2	$F_3 = \frac{s_3^2}{s_4^2}$	F_{p,f_3,f_4}
Reziduală	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{ij.})^2$	$f_4 = pq(n-1)$	s_4^2		
Totală	$\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{n}(x_{ijk} - \bar{x}_{...})^2$	$f = npq - 1$			

Scanned with OKEN Scanner

Se compară apoi F_c cu F_t şi, în funcţie de rezultatul obţinut pe baza unui anumit nivel de semnificaţie, se admite sau se respinge ipoteza nulă.

De exemplu, să presupunem că se studiază influenţa exercitată asupra recoltei de grâu (factorul $A: i = 1,2,\ldots,i,\ldots,p$) cărora li se administrează îngrăşăminte chimice de concentraţii diferite (factorul $B: j = 1,2,\ldots,j,\ldots,q$), cultivată pe npq parcele. Fiecare combinaţie i,j soi-îngrăşământ se cultivă pe n parcele.

Scanned with OKEN Scanner

13. LEGĂTURILE DINTRE FENOMENELE REALE

13.1. Noţiuni preliminare

Statistica folosindu-se de metodele proprii de cercetare, contribuie alături de celelalte ştiinţe la descoperirea legilor, legităţilor şi regularităţilor care se manifestă în interiorul şi între fenomenele colective. De altfel, problema fundamentală a oricărei cunoaşteri ştiinţifice, a oricărei ştiinţe este descoperirea legilor care acţionează asupra fenomenelor care fac obiectul lor de studiu.

Fenomenele şi procesele din natură şi societate iau naştere şi se dezvoltă sub influenţa acţiunii unui mare număr de factori principali şi secundari, esenţiali şi neesenţiali, cuantificabili şi necuantificabili sau cuantificabili cu aproximaţie.

În capitolele precedente ale acestei cărţi, au fost prezentate deja unele metode şi tehnici susceptibile de a facilita comparaţia între două sau mai multe fenomene colective cum ar fi: privirea comparativă a două serii paralele; compararea mediilor şi a gradului lor de împrăştiere, reprezentarea grafică, tabelul cu dublă intrare etc.

În această secţiune vom depăşi această primă fază descriptivă de studiere a legăturilor dintre fenomene, introducând elemente noi de măsurare mai precisă a lor.

Fenomenele naturale, care iau naştere ca urmare a mişcării mecanice, fizice, chimice şi biologice a materiei, sunt legate între ele prin *legături funcţionale*.

De exemplu, fiind dată o bară metalică supusă unor temperaturi diferite: $t_1, t_2, t_3, \ldots, t_n$ ea se va dilata şi va lua lungimile, respectiv: $l_1, l_2, l_3, \ldots, l_n$. De fiecare dată când bara va fi supusă temperaturii t_i, va lua lungimea l_i. Astfel, putem spune că unei temperaturi (dintr-un anumit interval al temperaturilor) îi corespunde (sau i se poate asocia) o anumită lungime a barei şi numai una. Legătura dintre temperatură şi lungimea barei este o legătură funcţională. Acest tip de legături, specific fenomenelor naturii, nu este aplicabil fenomenelor sociale şi economice.

Fenomenelor social-economice le este caracteristic faptul că asupra lor acţionează, alături de cauze determinate pentru legătura sau dependenţa cercetată, numeroşi alţi factori ne luaţi în calcul, accidentali, a căror acţiune poate influenţa dezvoltarea lor. Acest fapt determină existenţa unor legături de forme şi tipuri variate între fenomene cu grade diferite de complexitate. Natura legăturii dintre fenomene se stabileşte pe baza unei analize calitative multilaterale.

Scanned with OKEN Scanner

În unele cazuri, între fenomene există o *legătură cauzală*, atunci când unul dintre fenomene joacă rol de cauză, iar celălalt de efect. De exemplu, creşterea productivităţii muncii determină în mare măsură reducerea costului producţiei.

Între alte fenomene poate exista o *legătură sau o interacţiune reciprocă*, când nici unul dintre fenomene nu poate fi considerat cauză sau efect. Asfel, se constată cu caracter de legitate, că vârsta la căsătorie a soţului concordă cu cea a soţiei, deci se căsătoresc, în general, persoane care au o vârstă apropiată. În acest caz, nici vârsta soţului, nici cea a soţiei nu poate fi considerată numai cauză sau numai efect.

În alte cazuri, variaţia paralelă a două fenomene *nu se datorează legăturii dintre ele, ci influenţei unui al treilea factor*, care le determină pe ambele. Aşa, de exemplu, cererea de mărfuri şi economiile băneşti ale populaţiei variază în acelaşi sens, cu toate acestea nici unul dintre ele nu poate fi considerat cauză sau efect al celuilalt. Ambele fenomene sunt influenţate de un al treilea factor şi anume, de mărimea veniturilor băneşti ale populaţiei.

În cadrul fenomenelor sociale şi economice, *legătura nu se manifestă în fiecare caz în parte* - aşa cum se întâmplă în cadrul fenomenelor naturii - *ci în general şi în medie, într-un mare număr de cazuri.* Fenomenul efect este rezultatul conjugării influenţei mai multor fenomene-cauză şi în acest sistem de legături, nu toate raporturile de dependenţă au aceeaşi importanţă, iar uneori acţiunea unora dintre ele se compensează reciproc. Studierea acestor fenomene, cunoaşterea legilor, legităţilor şi regularităţilor de producere a lor nu se poate face pe cale experimentală, ci folosind mai ales metoda abstractizării, prin care se înlătură ceea ce este întâmplător şi neesenţial în dezvoltarea lor, desprinzând astfel raporturile obiective de cauzalitate şi descoperind condiţiile de formare şi dezvoltare a acestora.

Variaţiile cantitative ale caracteristicii (variabilei) efect Y pot fi mai mari sau mai mici decât variaţiile cantitative determinate de caracteristicile cauzale X_i explicite. Unele valori ale caracteristicii efect (sau rezultative) pot avea chiar sens contrar sensului de variaţie pe care-l au caracteristicile cauzale. O astfel de legătură între caracteristicile fenomenelor sau proceselor sociale colective poartă denumirea de *dependenţă* sau *legătură stohastică*.

Legăturile statistice sau *stohastice* se caracterizează prin faptul că o caracteristică X exercită o anumită influenţă asupra altei caracteristici Y, îi determină într-o anumită măsură variaţia, însă asupra caracteristicii Y mai acţionează şi alţi factori în afară de X, care, din punct de vedere al legăturii dintre X şi Y, se consideră întâmplători.

De exemplu, între gradul înzestrării tehnice a muncii şi nivelul productivităţii muncii există o legătură bine determinată în sensul că, pe măsura creşterii nivelului înzestrării tehnice a muncii, creşte şi nivelul productivităţii muncii; dar productivitatea muncii este influenţată şi de alţi factori, în afară de înzestrarea

Scanned with OKEN Scanner

tehnică, cum ar fi: experienţa în producţie, nivelul de calificare, salarizarea, forma de organizare, factori psihologici etc.

Legătura statistică dintre două sau mai multe caracteristici exprimate numeric poartă denumirea de corelaţie. În general, se poate spune că între două sau mai multe fenomene există corelaţie, dacă între ele există o anumită legătură logică, cuantificabilă.

În opoziţie cu legăturile funcţionale, se pot întâlni în realitate şi fenomene între care să nu existe nici un fel de legătură, adică fenomenele respective sunt independente între ele. Deci, corelaţia este intermediară între independenţa absolută dintre fenomene şi legăturile funcţionale.

Studiul corelaţiei este îndreptat spre *stabilirea existenţei unei legături reale* între fenomenele cercetate şi apoi, cu ajutorul unor metode şi tehnici adecvate, se *caracterizează direcţia, forma şi intensitatea legăturii.*

Mulţimea legăturilor statistice dintre fenomenele colective se poate clasifica după mai multe criterii cum ar fi: numărul variabilelor corelate, modul de exprimare a caracteristicii, direcţia legăturii, forma legăturii şi intensitatea ei.

1) După *numărul caracteristicilor corelate*, legăturile dintre fenomene se pot împărţi în două grupe: *legături simple şi legături multiple. Legăturile simple* se stabilesc numai între două caracteristici, în timp ce legăturile dintre trei sau mai multe caracteristici poartă denumirea de *legături multiple.* O astfel de legătură există (de exemplu) între recolta medie la hectar, pe de o parte şi calitatea solului, măsurile agrotehnice, condiţiile meteorologice, pe de altă parte; legătura dintre producţia industrială pe judeţe, capitalul fix şi numărul de personal din industrie etc.

Dacă legăturile dintre fenomene sunt cauzale, atunci caracteristicile care joacă rol de cauză poartă denumirea de *variabile* sau *caracteristici independente, factoriale* sau *determinante* şi se notează cu $X_1, X_2, X_3, \ldots, X_n$, iar caracteristica efect se mai numeşte şi *caracteristică* sau *variabilă determinată* sau *variabilă dependentă* sau *variabilă rezultativă* şi se notează cu Y.

2) După *direcţia sau sensul lor*, legăturile statistice pot fi *directe* sau *pozitive* şi *inverse* sau *negative.* Legăturile directe apar atunci când variabila (sau variabilele) factorială, cât şi variabila rezultativă variază mereu în acelaşi sens. Dacă variaţia caracteristicilor factoriale are sens contrar variaţiei caracteristicii rezultative, atunci legătura dintre ele este inversă. De exemplu, între înzestrarea tehnică a muncii şi nivelul productivităţii muncii există o legătură directă, iar între productivitatea muncii şi costul producţiei există o legătură inversă.

3) După *forma lor*, legăturile dintre variabile pot fi *liniare* sau *curbilinii* (hiperbolă, parabolă, exponenţială etc.).

Orice studiu statistic concret de corelaţie trebuie să pornească de la identificarea existenţei unei legături reale între fenomene şi apoi să treacă la măsurarea ei statistică. De asemenea, trebuie să se ţină seama de faptul că orice studiu concret se sprijină pe un număr limitat de observaţii şi că, pe baza rezultatelor

Scanned with OKEN Scanner

obţinute prin prelucrarea şi analiza multilaterală a acestora, se încearcă apoi caracterizarea colectivităţii generale. De aceea, rezultatele calculelor de regresie şi corelaţie trebuie să fie interpretate probabilistic.

Pentru studierea legăturilor statistice dintre fenomene se pot utiliza în practică *metode elementare* (metoda seriilor paralele interdependente, metoda grupărilor, metoda tabelului de corelaţie, metoda balanţelor, metoda grafică etc.) şi *metode analitice* (regresie şi corelaţie simplă şi multiplă, determinaţia simplă şi multiplă, verificarea ipotezei existenţei şi formei legăturii, verificarea semnificaţiei indicatorilor de regresie, corelaţie şi determinaţie).

13.2. Metode elementare de studiere a legăturilor dintre fenomene

Metodele statistice elementare de studiere a dependenţelor dintre fenomene sunt uşor de aplicat, se bazează pe analiza calitativă a variabilelor corelate şi oferă informaţii importante asupra naturii caracteristicilor legăturii cercetate. Ele servesc ca punct de plecare în vederea aplicării metodelor analitice mai complexe şi mai precise.

13.2.1. Metoda seriilor statistice interdependente

Această metodă constă în stabilirea existenţei şi caracterizarea legăturilor dintre fenomene pe baza comparării indicatorilor proveniţi din două serii paralele: una care prezintă variaţia variabilei factoriale X şi alta pentru variabila rezultativă Y. Se pot supune comparaţiei două serii de timp, două serii de spaţiu sau două serii de distribuţie.

Tabelul 13.1. Două serii paralele interdependente

X	Y
x_1	y_1
x_2	y_2
x_3	y_3
⋮	⋮
x_i	y_i
⋮	⋮
x_n	y_n
$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$

Scanned with OKEN Scanner

În primul caz, termenii celor două serii dinamice se înscriu în ordine cronologică, iar în cazul al doilea şi al treilea, se procedează la ordonarea termenilor seriilor în ordine crescătoare sau descrescătoare a variabilei factoriale X şi în paralel se înscriu termenii corespunzători ai variabilei rezultative Y. Prin aplicarea acestei metode este posibil să se obţină o primă imagine asupra existenţei şi direcţiei legăturii dintre variabile, dar nu se poate aprecia forma şi intensitatea ei.

În cazul seriilor cronologice interdependente este necesar ca, pe lângă identificarea legăturilor dintre fenomene, să se precizeze şi perioada pentru care are loc legătura respectivă. Astfel, se pot evidenţia *legături sincrone* dintre fenomene, când variaţia variabilei factoriale determină concomitent variaţia variabilei rezultative şi *legături asincrone* în cazul în care variaţia variabilei factoriale se exercită asupra variabilei rezultative, după o anumită periodă de timp, deci cu un anumit decalaj.

Procedeul seriilor paralele interdependente se poate aplica atât în cazul legăturilor simple, cât şi în cazul legăturilor multiple când este necesar ca, pe baza analizei critice calitative, să se procedeze la ierarhizarea influenţei factorilor cauzali în funcţie de mărimea influenţei lor asupra variabilei rezultative. Dacă variabilele factoriale nu sunt independente între ele, în cazul corelaţiei multiple, apare *fenomenul de multicoliniaritate*, care trebuie diminuat cât mai mult posibil înainte de a se trece la calculul indicatorilor de regresie şi corelaţie multiplă. Există în literatura de specialitate metode de verificare a existenţei multicoliniarităţii şi de diminuare a efectelor ei în cazul corelaţiei liniare multiple. Eliminarea ei totală nu este posibilă.

Aplicarea metodei seriilor paralele interdependente devine mai dificilă în cazul seriilor formate dintr-un număr foarte mare de termeni şi implică în mare măsură subiectivismul cercetătorului. Pentru a înlătura acest neajuns, se poate utiliza metoda grupărilor.

13.2.2. Metoda grupărilor

Metoda grupărilor reprezintă o metodă de bază în analiza legăturilor dintre fenomene şi constă în împărţirea unităţilor statistice după variaţia variabilei factoriale X şi calcularea indicatorilor sintetici (mărimi medii, dispersii, mărimi relative de structură) ai caracteristicii rezultative Y, pentru fiecare grupă în care a fost împărţită colectivitatea. Această metodă oferă posibilitatea aprecierii existenţei legăturii şi sensului ei, dar, ca şi în cazul metodei anterioare, aprecierea formei şi intensităţii legăturii este cu totul subiectivă, depinzând în mare măsură de competenţa şi experienţa cercetătorului.

În cazul legăturilor simple se utilizează grupările simple, iar în cazul legăturilor multiple se vor utiliza grupările combinate.

Scanned with OKEN Scanner

O formă mai dezvoltată a acestei metode este *metoda tabelului de corelaţie*, care se aplică în practică atunci când numărul perechilor de valori ale variabilelor corelate este foarte mare.

Tabelul 13.2. Macheta unui tabel de corelaţie

X \ Y	y_1	...	y_j	...	y_p	Total
x_1	n_{11}	...	n_{1j}	...	n_{1p}	$n_{1.}$
x_2	n_{21}	...	n_{2j}	...	n_{2p}	$n_{2.}$
⋮	⋮	...	⋮	...	⋮	⋮
x_i	n_{i1}	...	n_{ij}	...	n_{jp}	$n_{i.}$
⋮	⋮	...	⋮	...	⋮	⋮
x_k	n_{k1}	...	n_{kj}	...	n_{qp}	$n_{k.}$
Total	$n_{.i}$	...	$n_{.k}$	...	$n_{.p}$	$n_{..}$

În cadrul tabelului de corelaţie gruparea se poate face pe variantele celor două variabile sau pe intervale de variaţie. La stabilirea intervalelor de variaţie ale celor două variabile este necesar să se considere pentru fiecare variabilă acelaşi număr de intervale închise şi egale de variaţie. Cu ajutorul metodei tabelului de corelaţie se poate aprecia existenţa legăturii dintre variabile, direcţia şi intensitatea ei. Astfel, dacă frecvenţele n_{ij} se dispersează relativ uniform în întregul spaţiu al tabelului, înseamnă că între variabile nu există nici un fel de legătură. Dimpotrivă, dacă frecvenţele se concentrează în jurul unei drepte sau a unei curbe de un anumit tip înseamnă că între variabile există o legătură a cărei formă este dată de dreapta sau curba respectivă. Cu cât gradul de concentrare a frecvenţelor în jurul dreptei sau curbei este mai mare, cu atât se poate aprecia că intensitatea legăturii dintre variabile este mai mare şi invers. Dacă alura dreptei sau curbei se îndreaptă dinspre valorile mici ale celor două variabile către valorile lor mai mari, atunci se poate spune că între variabile există o legătură directă şi invers. Această ipoteză este mai apropiată de realitate în cazul legăturilor de formă liniară. O expresie mai precisă a formei

Scanned with OKEN Scanner

legăturilor dintre variabile se va obţine prin studiul concomitent al variaţiei perechilor de valori x_i şi mediile condiţionate ale variabilei rezultative $\overline{y}_{x_i}$.

O formă particulară a metodei tabelului de corelaţie este *metoda balanţelor* cu varianta ei cea mai evoluată - *balanţa legăturilor input-output.*

13.2.3. Metoda grafică

Această metodă permite prezentarea într-o formă geometrică, expresivă şi vizuală a legăturii dintre două sau mai multe variabile.

În cazul a două variabile, metoda grafică se bazează pe sistemul de axe rectangulare, în care valorile variabilei factoriale se înscriu pe abscisă, iar cele ale variabilei rezultative pe ordonată. Perechile de valori corelate x_i, y_i, se reprezintă prin câte un punct în spaţiul *xoy*. Mulţimea acestor puncte dispersate în grafic oferă imaginea câmpului de corelaţie asemănătoare cu un "*nor de puncte*" - de unde derivă şi denumirea acestui procedeu.

După forma şi poziţia norului de puncte în plan se poate aprecia existenţa, forma, direcţia şi intensitatea legăturii dintre variabile. Concentrarea punctelor în jurul unei anumite drepte sau curbe ne va sugera faptul că între variabile există o legătură (fig.13.1. *A* şi *B*). De asemenea, se poate aprecia că nu există legătură între variabile dacă punctele empirice se concentrează în jurul unei drepte care este paralelă cu una dintre axe (fig.13.2. *B* şi *C*).

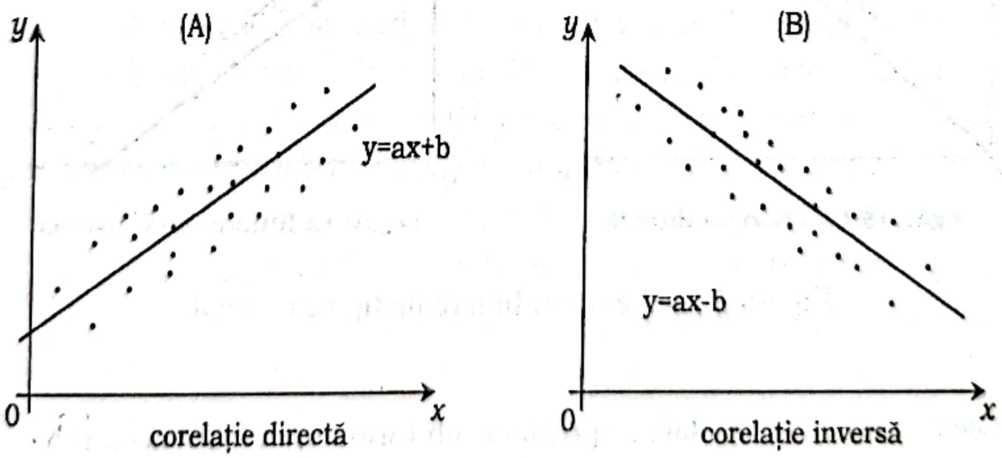


Figura 13.1. Tipuri de legături liniare

Scanned with OKEN Scanner

Dacă punctele empirice se situează chiar pe o dreaptă (fig.13.3. *A* şi *B*) sau pe o curbă de un anumit tip, atunci legătura dintre variabile este o legătură de tip funcţional.

În funcţie de modul de orientare a dreptei sau curbei care descrie forma legăturii dintre variabile în planul *xoy*, se poate aprecia şi direcţia legăturii dintre ele.

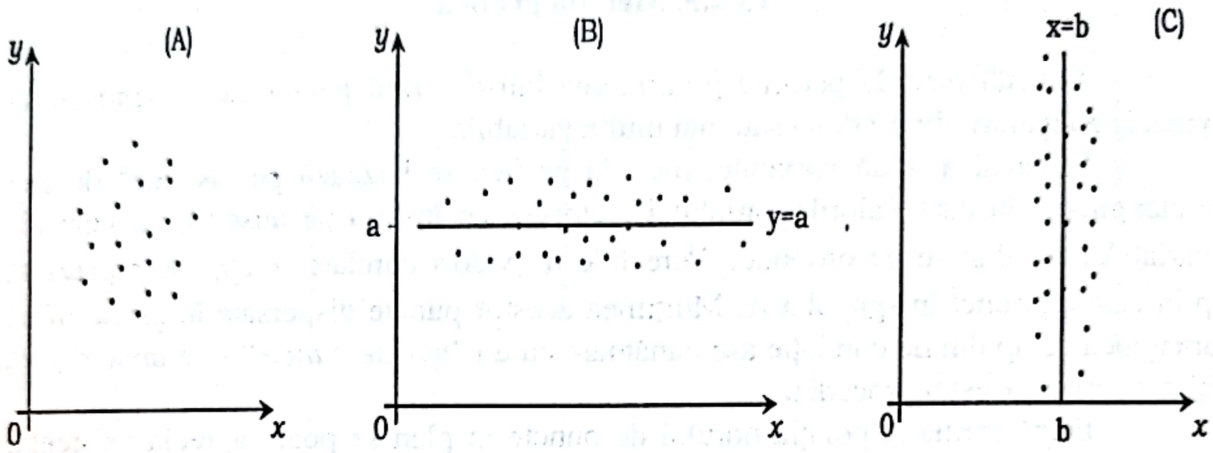


Figura 13.2. Independenţă liniară între variabile

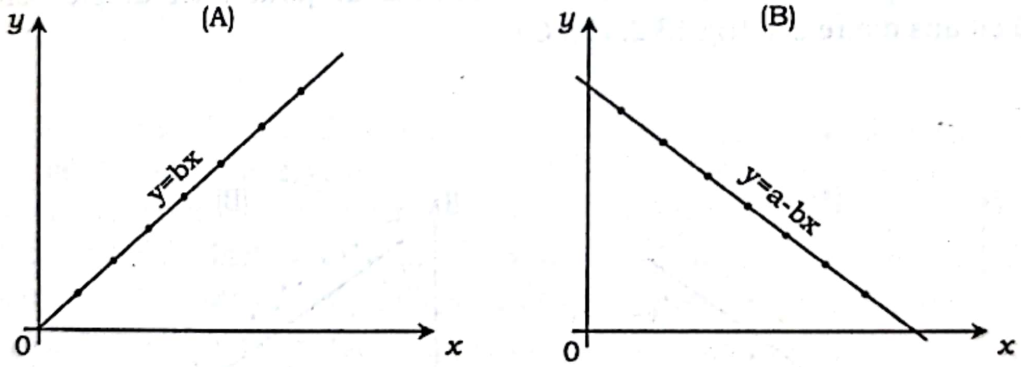


Figura 13.3. Legături liniare de tip funcţional

Dacă variabilele corelate se prezintă sub forma unui tabel de corelaţie, atunci pentru reprezentarea lor grafică se vor utiliza procedeele cunoscute: norul de puncte şi stereograma în plan şi diagrama prin paralelipipede dreptunghiulare ori suprafaţa poliedrală în spaţiul tridimensional.

Scanned with OKEN Scanner

14. REGRESIE ŞI CORELAŢIE SIMPLĂ LINIARĂ

14.1. Regresie simplă liniară

Cercetarea concretă a legăturii dintre fenomene numai pe baza metodelor elementare este insuficientă. Informaţiile obţinute cu acest prilej trebuie completate şi determinate cu mai mare precizie utilizând metode analitice. Cu ajutorul acestora se pot calcula indicatori sintetici privind expresia matematică a legăturii dintre fenomene, a formei, direcţiei şi intensităţii acesteia.

Cazul cel mai simplu de legătură dintre două variabile X şi Y este cel al regresiei liniare. Aceasta înseamnă că mulţimea punctelor x_i, y_i se distribuie geometric în jurul unei drepte :

$$y_x = a + bx \qquad (14.1)$$

Modelul regresiei liniare corespunde în practică legăturii dintre două variabile care variază proporţional.

Introdus în statistică de către statisticianul englez *Francis Galton* (1822-1911), termenul de *regresie* derivă de la latinescul "*regressio*" care înseamnă "*întoarcere*". În cadrul cercetărilor sale de ereditate, *F.Galton* a constatat că, din părinţi cu talie mai mică decât media pe colectivitate, se nasc copii cu o talie superioară mediei colectivităţii de copii şi invers. Deşi în cazul general denumirea de regresie este improprie, ea a devenit o noţiune consacrată în statistică.

Geometric, parametrii a şi b din ecuaţia (14.1) au semnificaţia din figura14.1. Parametrul a reprezintă *ordonata la origine* şi poate lua valori atât pozitive cât şi negative, iar parametrul b poartă denumirea de *coeficient de regresie* şi măsoară înclinaţia, panta sau coeficientul unghiular al dreptei faţă de axa absciselor: $b = tg\alpha$.

Determinarea parametrilor dreptei de regresie şi, în general, a oricărei funcţii de regresie se poate face cu ajutorul *metodei punctelor selecţionate* sau cu ajutorul *metodei celor mai mici pătrate*.

Scanned with OKEN Scanner

Calculul parametrilor funcţiei liniare cu ajutorul *metodei punctelor selecţionate* se poate efectua pe cale grafică, considerând două puncte situate pe dreapta trasată vizual la fiecare extremitate a segmentului

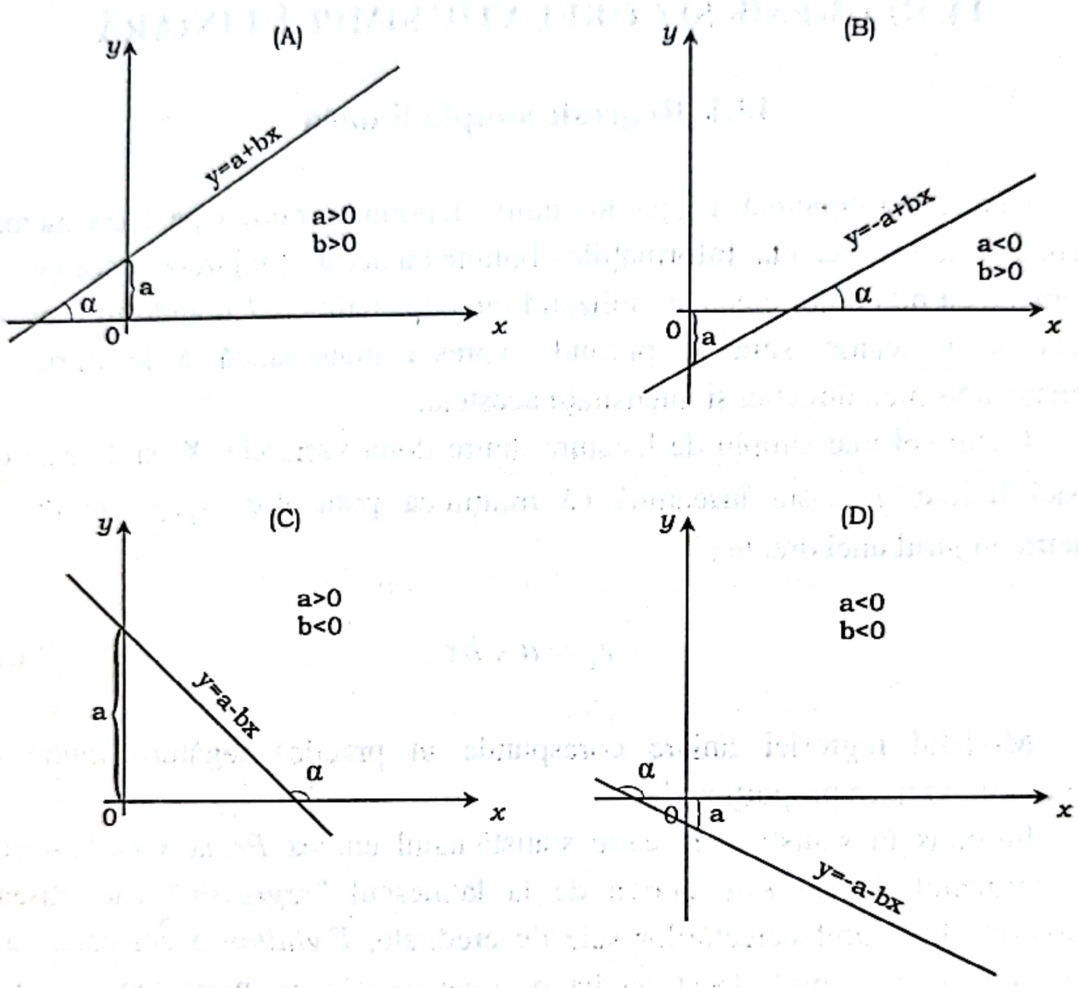


Figura 14.1. Tipuri de legături simple liniare

Să presupunem că s-au selecţionat punctele de coordonate x_1, y_1 şi x_2, y_2. Substituind aceste valori în ecuaţia (14.1), rezultă sistemul:

$$\begin{cases} a + bx_1 = y_1 \\ a + bx_2 = y_2 \end{cases} \qquad (14.2)$$

Rezolvând sistemul (14.2) se obţin valorile celor doi parametri a şi b.

Din geometria analitică se cunoaşte forma ecuaţiei unei drepte care trece prin două puncte x_1, y_1 şi x_2, y_2:

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1)$$

în care $\frac{y_2 - y_1}{x_2 - x_1}$ reprezintă coeficientul unghiular al dreptei.

Numărul punctelor selectate, în fiecare caz particular, este egal cu numărul parametrilor ce urmează a fi estimaţi şi deci egal cu numărul ecuaţiilor din sistem.

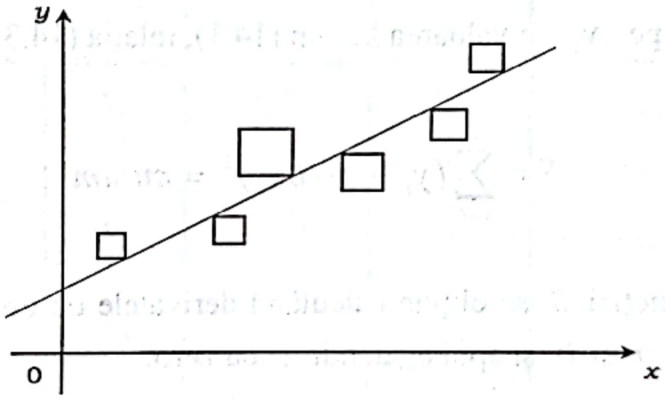


Figura 14.2. Cele mai mici pătrate ale abaterilor valorilor empirice faţă de valorile teoretice ale lui y

Trasarea vizuală a dreptei depinde de experienţa şi de competenţa cercetătorului şi poate să difere de la un cercetător la altul, ceea ce înseamnă că şi punctele care se selectează de cercetători diferiţi vor fi diferite, ei vor fi influenţaţi într-o anumită măsură de subiectivism. Subiectivismul cercetătorului ar putea fi diminuat prin împărţirea valorilor variabilelor corelate într-un număr de părţi egale (egal cu numărul parametrilor ecuaţiei de regresie) şi prin înlocuirea valorilor individuale ale coordonatelor punctelor situate pe dreapta trasată vizual cu mediile grupelor de termeni.

Scanned with OKEN Scanner

Caracterul subiectiv al dreptei de regresie, obţinute pe baza metodei punctelor selecţionate se înlătură în totalitate prin utilizarea *metodei celor mai mici pătrate MCMMP*.

Determinarea parametrilor dreptei de regresie cu ajutorul *MCMMP*, se bazează pe următorul principiu: *suma pătratelor diferenţelor dintre valorile empirice* y_i şi *valorile teoretice* y_{x_i}, date de ecuaţia de regresie, să fie minimă. Geometric, pătratul diferenţelor dintre y_i şi y_{x_i} se exprimă prin aria unor pătrate ale căror laturi sunt egale cu $y_i - y_{x_i}$ (fig.14.2). Este clar că suma ariilor acestor pătrate va fi cu atât mai mică, cu cât valorile empirice ale lui y se vor apropia mai mult de valorile teoretice corespunzătoare şi invers:

$$S = \sum_{i=1}^{n} (y_i - y_{x_i})^2 = minim \tag{14.3}$$

Substituind pe y_{x_i} cu valoarea lui din (14.1), relaţia (14.3) devine:

$$S = \sum_{i=1}^{n} (y_i - a - bx_i)^2 = minim \tag{14.4}$$

Minimul funcţiei S se obţine calculând derivatele de ordinul întâi în raport cu cei doi parametri a şi b şi apoi egalându-le cu zero:

$$\begin{cases} \dfrac{\partial S}{\partial a} = -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \dfrac{\partial S}{\partial b} = -2\sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0 \end{cases} \tag{14.5}$$

Prin transformări elementare din relaţiile (14.5), se obţine următorul sistem de ecuaţii, numit *sistem de ecuaţii normale*:

Scanned with OKEN Scanner

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{14.6}$$

Rezolvarea sistemului de ecuaţii normale (14.6) se poate efectua prin diferite metode, cum ar fi: *metoda substituţiei*, *metoda lui Cramer*, *metoda lui Gauss* şi altele.

Pentru determinarea parametrilor ecuaţiei de regresie simplă liniară pe baza sistemului (14.6), calculele intermediare se pot sistematiza sub forma tabelului 14.1.

Tabelul 14.1. Calculul parametrilor ecuaţiei de regresie simplă liniară

x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$y_{x_i} = a + bx_i$
x_1	y_1	x_1^2	$x_1 y_1$	y_1^2	$y_{x_1} = a + bx_1$
x_2	y_2	x_2^2	$x_2 y_2$	y_2^2	$y_{x_2} = a + bx_2$
x_3	y_3	x_3^2	$x_3 y_3$	y_3^2	$y_{x_3} = a + bx_3$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_i	y_i	x_i^2	$x_i y_i$	y_i^2	$y_{x_i} = a + bx_i$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_n	y_n	x_n^2	$x_n y_n$	y_n^2	$y_{x_n} = a + bx_n$
$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i y_i$	$\sum_{i=1}^{n} y_i^2$	$\sum_{i=1}^{n} y_{x_i} = a + b\sum_{i=1}^{n} x_i$

14.1.1. Metode de rezolvare a sistemelor de ecuaţii normale liniare

1) *Metoda substituţiei.* Din prima ecuaţie a sistemului (14.6), rezultă:

$$a = \frac{\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i}{n}$$

$$\bar{y} = a + b\bar{x} \qquad (14.7)$$

Relaţia (14.7) exprimă faptul că dreapta de regresie trece prin punctul de coordonate $(\bar{x}, \bar{y}\,1)$.

Pe baza relaţiei (14.7), ecuaţia dreptei (14.1) devine:

$$y_x = \bar{y} - b\bar{x} + bx_i$$

sau

$$y_x = \bar{y} + b(x_i - \bar{x}) \qquad (14.8)$$

Relaţia (14.8) exprimă valorile teoretice ale variabilei dependente Y în funcţie de coeficientul de regresie b şi de mediile $\bar{x}$ şi $\bar{y}$ ale celor două variabile, numit *centru de greutate* al norului de puncte.

Substituind pe a din (14.7) în ecuaţia a doua a sistemului (14.6), rezultă:

$$\left(\frac{\sum_{i=1}^{n} y_i}{n} - b \frac{\sum_{i=1}^{n} x_i}{n} \right) \sum_{i=1}^{n} x_i + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

de unde

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2}$$

sau

$$b = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2} \qquad (14.9)$$

Împărţind numărătorul şi numitorul relaţiei (14.9) la n rezultă:

$$b = \frac{\frac{\sum_{i=1}^{n} x_i y_i}{n} - \overline{x}\,\overline{y}}{\frac{\sum_{i=1}^{n} x_i^2}{n} - \overline{x}^2} = \frac{cov(xy)}{\sigma_x^2}$$

deci $$b = \frac{cov(xy)}{\sigma_x^2} \qquad (14.10)$$

Dacă se face o mişcare de translaţie a dreptei de regresie în aşa fel, încât originea axelor de coordonate să coincidă cu punctul de coordonate $\overline{x}$ şi $\overline{y}$, valoarea coeficientului de regresie b nu se modifică, deoarece înclinaţia dreptelor faţă de axa absciselor rămâne neschimbată.

Astfel, substituind $x_i' = x_i - \overline{x}$ şi $y_i' = y_i - \overline{y}$ în (14.7) rezultă:

$$a = \frac{\sum_{i=1}^{n}(y_i - \overline{y}) - b\sum_{i=1}^{n}(x_i - \overline{x})}{n} = 0$$

sau

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (14.11)$$

În urma mişcării de translaţie a dreptei de regresie în aşa fel încât originea axelor de coordonate să coincidă cu punctul de abscisă $\bar{x}$, atunci sistemul de ecuaţii normale (14.6) devine:

$$na + b\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} y_i$$

$$a\sum_{i=1}^{n}(x_i - \bar{x}) + b\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})y_i \qquad (14.12)$$

Deoarece $\sum_{i}(x_i - \bar{x}) = 0$, sistemul (14.12) se poate scrie astfel:

$$na = \sum_{i=1}^{n} y_i$$

$$b\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i - \bar{x})y_i$$

Iar formulele uzuale de calcul ale parametrilor ecuaţiei regresiei simple liniare vor avea expresiile:

$$a = \bar{y} \qquad (14.13)$$

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (14.14)$$

Scanned with OKEN Scanner

Să considerăm, pentru exemplificare, legătura dintre producţia de energie termoelectrică şi producţia de cărbune extras în *România* în perioada 1960 -1976 (tabelul 14.2).

Tabelul 14.2. Producţia de cărbune extras şi de energie termoelectrică a *României* în perioada 1960-1976

Anul	Cărbune extras (milioane tone)	Energie termoelectrică
1960	8,2	7,3
1961	8,7	8,2
1962	9,6	9,4
1963	10,3	11,1
1964	11,1	13,3
1965	12,1	16,2
1966	13,5	19,8
1967	15,0	23,3
1968	17,0	26,3
1969	19,2	29,3
1970	22,8	32,3
1971	23,0	35,0
1972	25,3	36,1
1973	26,7	39,2
1974	29,2	40,6
1975	29,4	45,0
1976	28,1	50,2

Sursa: *Anuarele statistice ale României 1961-1977*, DCS, Bucureşti

Legătura dintre aceste două variabile este evidentă. Cea mai mare parte din producţia de cărbune este utilizată în termocentrale pentru producţia de energie termoelectrică, dar cărbunele nu este singurul combustibil utilizat în termocentrale.

Legătura dintre producţia de energie termoelectrică (variabilă rezultativă) şi cantitatea de cărbune extras (variabilă independentă) se prezintă sugestiv în figura 14.3.

Pentru aplicarea formulelor (14.13) şi (14.14), în vederea determinării parametrilor ecuaţiei liniare de regresie, calculele se sistematizează cu ajutorul tabelului 14.3.

Scanned with OKEN Scanner

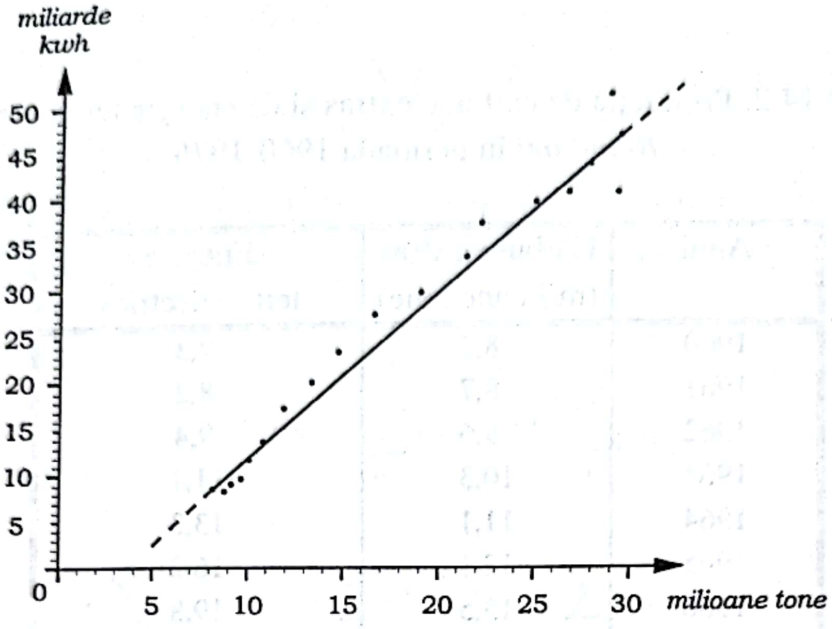


Figura 14.3. Legătura dintre producţia de cărbune şi cea de energie termoelectrică

Substituind datele din tabelul 14.2 în formulele (14.13) şi (14.14), rezultă:

$$a = \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{442,6}{17} = 26,35 \ \textit{miliarde kwh}$$

iar

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) y_i}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{1677,205}{961,318} = 1,744$$

Scanned with OKEN Scanner

Înlocuind succesiv valorile parametrilor a şi b, de mai sus, în ecuaţia de regresie (14.8) se obţin valorile teoretice ale variabilei dependente y_{x_i} (coloana 6 a tabelului 14.3).

Tabelul 14.3. Determinarea parametrilor ecuaţiei liniare de regresie prin metoda substituţiei

Anul	Cărbune extras (mil.t) x_i	Energie termo-electrică (miliarde kwh) y_i	$x_i - \bar{x}$	$(x_i - \bar{x})y_i$	$(x_i - \bar{x})^2$	y_x
0	1	2	3	4	5	6
1960	8,2	7,3	-9,988	-72,912	99,760	8,616
1961	8,7	8,2	-9,488	-77,802	90,022	9,488
1962	9,6	9,4	-8,588	-80,727	73,754	11,058
1963	10,3	11,1	-7,888	-87,557	62,221	12,278
1964	11,1	13,3	-7,088	-94,270	50,240	13,674
1965	12,1	16,2	-6,088	-98,626	37,064	15,418
1966	13,5	19,8	-4,688	-92,822	21,977	17,859
1967	15,0	23,3	-3,188	-74,280	10,163	20,475
1968	17,0	26,3	-1,188	-31,244	1,411	23,963
1969	19,2	29,3	-1,012	29,652	1,024	27,800
1970	22,8	32,3	4,612	148,968	21,271	34,078
1971	23,0	35,0	4,812	168,420	23,155	34,427
1972	25,3	36,1	7,112	256,743	50,581	38,438
1973	26,7	39,2	8,512	333,670	72,454	40,880
1974	29,2	40,6	11,012	447,087	121,264	45,240
1975	29,4	45,0	11,212	504,540	125,709	45,589
1976	28,1	50,2	9,912	497,582	98,248	43,322
Total	309,2	442,6	0	1677,205	961,318	442,603

Valorile teoretice ale variabilei dependente s-au reprezentat în figura14.3, de unde se observă că dreapta de regresie se apropie în mare măsură de valorile

Scanned with OKEN Scanner

empirice, ceea ce înseamnă că modelul liniar de regresie descrie în mod corespunzător legătura dintre cele două variabile.

Verificarea exactităţii calculelor ecuaţiei liniare de regresie se poate efectua pe baza relaţiei care există între valorile reale ale variabilei dependente y_i şi valorile teoretice y_{x_i} ale acestei variabile şi anume, *suma valorilor variabilei empirice este egală cu suma valorilor ei teoretice*, adică:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_{x_i}$$

În exemplul nostru, totalul pe coloana 2 a tabelului 14.3 este aproximativ egal cu totalul de pe coloana 6 a aceluiaşi tabel. Diferenţa de 0,3‰ dintre aceste sume se datorează numărului limitat de zecimale luate în calcul.

2) *Metoda lui Cramer.* Aplicarea acestei metode presupune calculul şi analiza prealabilă a determinantului principal şi a determinanţilor caracteristici corespunzători celor doi parametri.

Valoarea determinantului principal, obţinut pe baza sistemului (14.6), se determină astfel:

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = \left[n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2 \right] > 0$$

Valorile determinanţilor caracteristici vor fi:

$$\Delta_a = \begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i$$

Scanned with OKEN Scanner

$$\Delta_b = \begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i \end{vmatrix} = n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i$$

Aplicând regula lui *Cramer*, pentru rezolvarea sistemului de ecuaţii normale (14.6), rezultă următoarele formule uzuale pentru parametrii ecuaţiei de regresie:

$$a = \frac{\Delta_a}{\Delta} = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad (14.15)$$

$$b = \frac{\Delta_b}{\Delta} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} \qquad (14.16)$$

Pentru exemplificare, considerăm datele din tabelul 14.1. Calculele intermediare pentru determinarea parametrilor ecuaţiei regresiei simple liniare pe baza formulelor (14.15) şi (14.16) se prezintă sistematizat în tabelul 14.3.

Înlocuind în formulele (14.15) şi (14.16) elementele de calcul din tabelul 14.3, rezultă:

$$a = \frac{6584{,}12 \cdot 442{,}6 - 309{,}2 \cdot 9726{,}46}{17 \cdot 658{,}12 - 309{,}2^2} = \frac{-93289{,}9}{16325{,}4} = -5{,}7144$$

şi

$$b = \frac{17 \cdot 9726{,}46 - 309{,}2 \cdot 442{,}6}{17 \cdot 658{,}12 - 309{,}2^2} = \frac{28497{,}9}{16325{,}4} = 1{,}7456$$

Substituind valorile parametrilor a şi b în formula ecuaţiei de regresie rezultă valorile teoretice ale variabilei dependente y_{x_i} (coloana 6 din tabelul 14.3):

Scanned with OKEN Scanner

$$y_{x_i} = -5{,}7144 + 1{,}7456 \cdot x_i$$

Tabelul 14.4. Calculul coeficienţilor parametrilor ecuaţiei liniare de regresie pe baza metodei lui *Cramer*

Anul	x_i	y_i	x_i^2	y_i^2	$x_i y_i$	y_x	x_y
0	1	2	3	4	5	6	7
1960	8,2	7,3	67,24	53,29	59,86	8,599	7,837
1961	8,7	8,2	75,69	67,24	71,34	9,972	8,333
1962	9,6	9,4	92,16	88,36	90,24	11,043	8,996
1963	10,3	11,1	106,09	123,21	114,33	12,265	9,934
1964	11,1	13,3	123,21	176,89	147,63	13,662	11,149
1965	12,1	16,2	146,41	262,44	196,02	15,407	12,749
1966	13,5	19,8	182,25	392,04	267,30	17,851	14,737
1967	15,0	23,3	225,00	542,89	349,50	20,470	16,669
1968	17,0	26,3	289,00	691,69	447,10	23,961	18,325
1969	19,2	29,3	368,64	858,49	562,56	27,801	19,981
1970	22,8	32,3	519,84	1043,29	736,44	34,085	21,637
1971	23,0	35,0	529,00	1225,00	805,00	34,434	23,127
1972	25,3	36,1	640,09	1303,21	913,33	38,449	23,734
1973	26,7	39,2	712,89	1536,64	1046,64	40,893	25,445
1974	29,2	40,6	852,64	1648,36	1185,52	45,257	26,218
1975	29,4	45,0	864,36	2025,00	1323,00	45,606	28,647
1976	28,1	50,2	789,61	2520,04	1410,62	43,337	31,517
Total	309,2	442,6	6584,12	14558,08	9726,46	442,592	309,035

Ca şi în cazul metodei substituţiei, se verifică şi acum relaţia:

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_{x_i}$$

Coeficientul de regresie b = 1,7456 are o valoare apropiată de cea obţinută pe baza rezolvării sistemului de ecuaţii normale cu ajutorul metodei substituţiei şi arată că între cantitatea de cărbune extras şi producţia de energie termoelectrică există o legătură directă şi anume, la o creştere a producţiei de cărbune cu un milion tone, producţia de energie electrică creşte cu 1,74 miliarde kwh.

3) *Metoda lui Gauss* se bazează pe o succesiune de transformări elementare asupra unei matrici pătratice în aşa fel, încât matricea iniţială se transformă într-o

Scanned with OKEN Scanner

matrice unitate. Aceasta înseamnă că transformarea matricii $A_{n\times n}$ într-o matrice $I_{n\times n}$ se poate obţine cu ajutorul unor succesiuni de transformări liniare care dau acelaşi rezultat, ca şi înmulţirea matricii pătratice cu inversa ei:

$$A^{-1}_{n\times n} \cdot A_{n\times n} = I_{n\times n} \tag{14.17}$$

Invers, dacă aceleaşi transformări se aplică unei matrici unitate se va obţine matricea inversă $A^{-1}_{n\times n}$.

Pe baza acestui principiu, se pot rezolva sisteme de ecuaţii în care numărul ecuaţiilor este egal cu numărul necunoscutelor:

$$A_{n\times n} \cdot X_{n\times 1} = P_{n\times 1} \tag{14.18}$$

Soluţia X^* se obţine formând matricea extinsă $(A\ P)$ şi efectuând transformări liniare care vor reduce partea din stânga $A_{n\times n}$ a acestei matrici la o matrice unitate $I_{n\times n}$. În acelaşi timp, aceste transformări se vor aplica şi asupra părţii din dreapta $P_{n\times 1}$, deoarece transformările liniare vor fi aplicate asupra tuturor celor $n+1$ elemente de pe fiecare linie. Rezultatul acestor transformări $(A\ P) \Rightarrow (I\ X^*)$ este acelaşi, ca şi cum am fi înmulţit P cu matricea inversă A^{-1}, adică $A^{-1}\ P = X^*$.

Să aplicăm această metodă pentru rezolvarea sistemului de ecuaţii normale (14.6):

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

Matricea extinsă va fi:

$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

Scanned with OKEN Scanner

Se urmăreşte reducerea primelor două coloane ale acestei matrici la o matrice unitate $(2\,x\,2)$ prin transformări pe linii. O metodă sistematică necesită transformarea la început a elementului $(1,1)$ în 1, iar celelalte elemente de pe aceeaşi coloană în zero. Apoi se transformă următorul element de pe diagonala principală $(2,2)$ în unitate şi celelalte elemente de pe coloana 2 în zero. În continuare, transformările se vor aplica asupra elementului $(3,3)$ şi aşa mai departe, până când elementul (n,n) devine egal cu 1 şi celelalte elemente de pe coloana n devin zero.

În cazul nostru, $n = 2$, procedeul se va repeta numai de două ori. Schematic, succesiunea transformărilor şi a rezultatelor în fiecare treaptă se prezintă astfel:

Matricea extinsă
$$\begin{pmatrix} n & \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i \end{pmatrix}$$

Treapta I: se înmulţeşte linia 1 cu $\frac{1}{n}$:
$$\begin{pmatrix} 1 & \frac{\sum_{i=1}^{n} x_i}{n} & \frac{\sum_{i=1}^{n} y_i}{n} \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x\ y_i \end{pmatrix}$$

Treapta II: se adună la linia 2 scalarul $(-\sum_{i=1}^{n} x_i)$ înmulţit cu linia 1:
$$\begin{pmatrix} 1 & \frac{\sum_{i=1}^{n} x_i}{n} & \frac{\sum_{i=1}^{n} y_i}{n} \\ 0 & \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n} & \sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n} \end{pmatrix}$$

Scanned with OKEN Scanner

Treapta III : se înmulţeşte linia 2 cu scalarul $\dfrac{1}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ şi se simplifică expresia folosind notaţiile pentru medii şi momente simple:

$$\begin{pmatrix} 1 & \bar{x} & \bar{y} \\ 0 & 1 & \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \end{pmatrix}$$

Treapta IV: se aduna la linia 1 scalarul $(-\bar{x})$ înmulţit cu linia a doua:

$$\begin{pmatrix} 1 & 0 & \bar{y}-\dfrac{\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \\ 0 & 1 & \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \end{pmatrix}$$

Deci, sistemul de ecuaţii normale iniţial s-a transformat într-un sistem echivalent:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \bar{y}-b\bar{x} \\ \dfrac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \end{pmatrix}$$

Scanned with OKEN Scanner

De unde rezultă:

$$a = \bar{y} - b\bar{x}$$

şi

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Menţionăm că aceste formule sunt identice cu cele obţinute anterior prin aplicarea metodei substituţiei (vezi 14.7 şi 14.11).

Tabelul 14.5. Metoda lui *Gauss* de rezolvare a sistemelor de ecuaţii liniare

	a	*b*	
a *b*	17 309,2	309,2 6584,12	442,6 9726,46
a *b*	1 0	18,188235 961,318	26,035294 1676,3471
a *b*	1 0	0 1	-5,714 1,744

Pentru ilustrarea metodei lui *Gauss* de rezolvare a unui sistem liniar, să substituim în (14.6) elementele de calcul din tabelul 14.4, astfel:

$$17\,a + 309{,}2\,b = 442{,}6$$
$$309{,}2\,a + 6584{,}12\,b = 9726{,}46$$

Din ultima iteraţie rezultă:

$$a = -5{,}714$$

$$b = 1{,}744$$

Deci, ecuaţia de regresie care descrie legătura dintre producţia de energie electrică şi cantitatea de cărbune extras va fi de forma:

Scanned with OKEN Scanner

$$y_x = -5{,}714 + 1{,}744x_i$$

Dacă perechile de valori x_i, y_i se întâlnesc în cadrul distribuţiei de mai multe ori n_i, atunci cele două variabile vor apare ponderate în cadrul sistemului de ecuaţii normale de regresie. În acest caz, aplicarea metodei celor mai mici pătrate presupune:

$$S = \sum_{i=1}^{p} n_i (y_i - y_{x_i})^2 = minim \tag{14.19}$$

$$S = \sum_{i=1}^{p} n_i (y_i - a - bx_i)^2 = minim$$

Diferenţiind funcţia S în raport cu fiecare din cei doi parametri, se obţine:

$$\begin{cases} \sum_{i=1}^{p} n_i (y_i - a - bx_i) = 0 \\ \sum_{i=1}^{p} n_i x_i (y_i - a - bx_i) = 0 \end{cases}$$

iar prin transformări elementare rezultă:

$$\begin{cases} a\sum_{i=1}^{p} n_i + b\sum_{i=1}^{p} x_i n_i = \sum_{i=1}^{p} y_i n_i \\ a\sum_{i=1}^{p} x_i n_i + b\sum_{i=1}^{p} x_i^2 n_i = \sum_{i=1}^{p} x_i y_i n_i \end{cases} \tag{14.20}$$

Rezolvând sistemul (14.20), se obţin formulele de calcul ale celor doi parametri:

$$a = \frac{\sum_{i=1}^{p} x_i^2 n_i \sum_{i=1}^{p} y_i n_i - \sum_{i=1}^{p} x_i n_i \sum_{i=1}^{p} x_i y_i n_i}{\sum_{i=1}^{p} n_i \sum_{i=1}^{p} x_i^2 n_i - (\sum_{i=1}^{p} x_i n_i)^2} \tag{14.21}$$

Scanned with OKEN Scanner

$$b = \frac{\sum_{i=1}^{p} n_i \sum_{i=1}^{p} x_i y_i n_i - \sum_{i=1}^{p} x_i n_i \sum_{i=1}^{p} y_i n_i}{\sum_{i=1}^{p} n_i \sum_{i=1}^{p} x_i^2 n_i - (\sum_{i=1}^{p} x_i n_i)^2} \qquad (14.22)$$

Pentru determinarea practică a parametrilor ecuaţiei de regresie y_x, calculele se pot sintetiza sub forma tabelului 14.6.

Tabelul 14.6. Schema calculelor intermediare ale parametrilor ecuaţiei de regresie

Nr.crt	x_i	y_i	n_i	$x_i n_i$	$y_i n_i$	$x_i^2 n_i$	$x_i y_i n_i$	$y_i^2 n_i$
1	x_1	y_1	n_1	$x_1 n_1$	$y_1 n_1$	$x_1^2 n_1$	$x_1 y_1 n_1$	$y_1^2 n_1$
2	x_2	y_2	n_2	$x_2 n_2$	$y_2 n_2$	$x_2^2 n_2$	$x_2 y_2 n_2$	$y_2^2 n_2$
3	x_3	y_3	n_3	$x_3 n_3$	$y_3 n_3$	$x_3^2 n_3$	$x_3 y_3 n_3$	$y_3^2 n_3$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	x_i	y_i	n_i	$x_i n_i$	$y_i n_i$	$x_i^2 n_i$	$x_i y_i n_i$	$y_i^2 n_i$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	x_k	y_k	n_k	$x_k n_k$	$y_k n_k$	$x_k^2 n_k$	$x_k y_k n_k$	$y_k^2 n_k$
Total	-	-	$\sum_i n_i$	$\sum_i x_i n_i$	$\sum_i y_i n_i$	$\sum_i x_i^2 n_i$	$\sum_i x_i y_i n_i$	$\sum_i y_i^2 n_i$

Aşa cum am văzut anterior, în cazul unui număr mare de valori ale celor două variabile, pentru a reduce volumul calculelor se poate recurge la gruparea variantelor caracteristicilor corelate pe intervale de variaţie. Gruparea, pe lângă avantajele ei incontestabile, prezintă, după cum se ştie şi dezavantajul că determină o anumită pierdere de informaţie. De aceea, valorile indicatorilor de regresie şi corelaţie, calculaţi pe baza datelor negrupate, diferă de cei care se obţin cu ajutorul datelor grupate. Diferenţa dintre rezultate se micşorează pe măsură ce distribuţiile empirice, obţinute pe baza grupării, se apropie tot mai mult de o distribuţie normală. Subliniem faptul că este discutabilă calcularea indicatorilor de regresie şi corelaţie pe baza unor date grupate pe intervale neegale de variaţie.

Dacă numărul perechilor de valori (x_i, y_i) este mare, atunci pentru simplificarea calculelor este preferabilă sistematizarea datelor iniţiale sub formă de tabel cu dublă intrare. Numărul variantelor, ca şi numărul intervalelor de variaţie ale

Scanned with OKEN Scanner

celor două variabile corelate cuprinse în tabel trebuie să fie acelaşi. În acest caz, sistemul de ecuaţii normale ia forma:

$$\begin{cases} a\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} + b\sum_{i=1}^{k} x_i n_{i.} = \sum_{j=1}^{s} y_j n_{.j} \\ a\sum_{i=1}^{k} x_i n_{i.} + b\sum_{i=1}^{k} x_i^2 n_{i.} = \sum_{i=1}^{k}\sum_{j=1}^{s} x_i y_j n_{ij} \end{cases} \qquad (14.23)$$

Împărţind ambii membri ai primei ecuaţii din sistemul (14.23) la $\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} = n$, rezultă:

$$a + b\bar{x} = \bar{y}$$

de unde se obţine $$a = \bar{y} - b\bar{x}$$

unde $$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_{i.}}{\sum_{i=1}^{k} n_{i.}} \quad \text{şi} \quad \bar{y} = \frac{\sum_{j=1}^{s} y_j n_{.j}}{\sum_{j-1}^{s} n_{.j}}$$

Rezolvând sistemul (14.23), se obţin formulele uzuale de calcul ale parametrilor ecuaţiei liniare de regresie:

$$a = \frac{\sum_{i=1}^{k} x_i^2 n_{i.} \sum_{j=1}^{s} y_j n_{.j} - \sum_{i=1}^{k} x_i n_{i.} \sum_{i=1}^{k}\sum_{j=1}^{s} x_i y_j n_{ij}}{\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} \sum_{i=1}^{k} x_i^2 n_{i.} - (\sum_{i=1}^{k} x_i n_{i.})^2} \qquad (14.24)$$

Scanned with OKEN Scanner

Tabelul 14.7. Calculele intermediare ale parametrilor ecuaţiei liniare de regresie în cazul unui tabel de corelaţie

X \ Y	y_1	y_1	...	y_j	...	y_s	Total $(n_{i\cdot})$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$\sum_{j=1}^{s} y_j x_i n_{ij}$	y_{x_i}
x_1	n_{11}	n_{12}	...	n_{1j}	...	n_{1s}	$n_{1\cdot}$	$x_1 n_{1\cdot}$	$x_1^2 n_{1\cdot}$	$\sum_{j=1}^{s} y_j x_1 n_{1j}$	y_{x_1}
x_2	n_{21}	n_{22}	...	n_{2j}	...	n_{2s}	$n_{2\cdot}$	$x_2 n_{2\cdot}$	$x_2^2 n_{2\cdot}$	$\sum_{j=1}^{s} y_j x_2 n_{2j}$	y_{x_2}
...	...	...	...	...	...	...	...	...	...	...	...
x_i	n_{i1}	n_{i2}	...	n_{ij}	...	n_{is}	$n_{i\cdot}$	$x_i n_{i\cdot}$	$x_i^2 n_{i\cdot}$	$\sum_{j=1}^{s} y_j x_i n_{ij}$	y_{x_i}
...	...	...	...	...	...	...	...	...	...	...	...
x_k	n_{k1}	n_{k2}	...	n_{kj}	...	n_{ks}	$n_{k\cdot}$	$x_k n_{k\cdot}$	$x_k^2 n_{k\cdot}$	$\sum_{j=1}^{s} y_j x_k n_{kj}$	y_{x_k}
Total $(n_{\cdot j})$	$n_{\cdot 1}$	$n_{\cdot 2}$	...	$n_{\cdot j}$	...	$n_{\cdot s}$	$n_{\cdot\cdot}$	$\sum_{i=1}^{k} x_i n_{i\cdot}$	$\sum_{i=1}^{k} x_i^2 n_{i\cdot}$	$\sum_{i=1}^{k}\sum_{j=1}^{s} y_j x_i n_{ij}$	-
$y_j n_{\cdot j}$	$y_1 n_{\cdot 1}$	$y_2 n_{\cdot 2}$	...	$y_j n_{\cdot j}$	...	$y_s n_{\cdot s}$	$\sum_{j=1}^{s} y_j n_{\cdot j}$				
$y_j^2 n_{\cdot j}$	$y_1^2 n_{\cdot 1}$	$y_2^2 n_{\cdot 2}$	...	$y_j^2 n_{\cdot j}$	...	$y_s^2 n_{\cdot s}$	$\sum_{j=1}^{s} y_j^2 n_{\cdot j}$				
$\sum_{i=1}^{k} x_i y_j n_{ij}$	$\sum_{i=1}^{k} x_i y_1 n_{i1}$	$\sum_{i=1}^{k} x_i y_2 n_{i2}$	...	$\sum_{i=1}^{k} x_i y_j n_{ij}$	...	$\sum_{i=1}^{k} x_i y_s n_{is}$	$\sum_{j=1}^{1}\sum_{i=1}^{k} x_i y_j n_{ij}$				
x_{y_j}	x_{y_1}	x_{y_2}	...	x_{y_j}	...	x_{y_s}	-				

Scanned with OKEN Scanner

$$b = \frac{\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} \sum_{i=1}^{k}\sum_{j=1}^{s} x_i y_j n_{ij} - \sum_{i=1}^{k} x_i n_{i.} \sum_{j=1}^{s} y_j n_{.j}}{\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} \sum_{i=1}^{k} x_i^2 n_{i.} - (\sum_{i=1}^{k} x_i n_{i.})^2} \qquad (14.25)$$

Împărţind ambii membri ai relaţiei (14.25) la $\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij} = n$, se obţine:

$$b = \frac{\sum_{i=1}^{k}\sum_{j=1}^{s} x_i y_j n_{ij} - \bar{x}\,\bar{y}}{\frac{1}{\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij}} \sum_{i=1}^{k} x_i^2 n_{i.} - \bar{x}^2} = \frac{cov(x,y)}{\sigma_x^2}$$

Să considerăm, pentru exemplificarea calculului parametrilor ecuaţiei de regresie liniară, în cazul sistematizării celor două variabile sub forma unui tabel cu dublă intrare, distribuţia judeţelor *României* după valoarea fondurilor fixe şi după numărul mediu al personalului. (tabelul 14.8).

Tabelul 14.8. Distribuţia judeţelor după valoarea fondurilor fixe şi numărul mediu al personalului în anul 1976

Miliarde lei fonduri fixe \ Mii persoane	-50	50-100	100-150	150-200	200 -	Total
- 10	3	3	-	-	-	6
10 - 20	-	8	3	-	-	11
20 - 30	-	4	6	2	-	12
30 - 40	-	-	2	1	1	4
40 -	-	-	-	3	3	6
Total	3	15	11	6	4	39

Sursa: *Anuarul statistic al României, 1977*, DCS, p.93 şi 104-105

Scanned with OKEN Scanner

Tabelul 14.9. Calculul parametrilor ecuaţiei regresiei simple liniare în cazul unei distribuţii bidimensionale X

Nr. crt	X \ Y	- 50	50 - 100	100 - 150	150 - 200	200 -	Total ($n_{i.}$)	x_i	$x_i n_{i.}$	$x_i^2 n_{i.}$	$\sum_{j=1}^{s} x_i y_j n_{ij}$	y_{x_i}
0	1	2	3	4	5	6	7	8	9	10	11	12
1	- 10	3	3	-	-	-	6	5	30	150	1500	48,59
2	10 - 20	-	8	3	-	-	11	15	165	2475	14625	85,63
3	20 - 30	-	4	6	2	-	12	25	300	7500	35000	122,67
4	30 - 40	-	-	2	1	1	4	35	140	4900	22750	159,72
5	40 -	-	-	-	3	3	6	45	270	12150	54000	196,76
6	Total ($n_{.j}$)	3	15	11	6	4	39	-	905	27175	127875	-
7	y_j	25	75	125	175	225	-					
8	$y_j n_{.j}$	75	1125	1375	1050	900	4525					
9	$y_j^2 n_{.j}$	1875	28125	103125	131250	202500	466875					
10	$\sum_{i=1}^{k} x_i y_j n_{ij}$	375	17625	33125	38500	38250	127875					

Scanned with OKEN Scanner

Calculele intermediare necesare determinării parametrilor ecuaţiei de regresie se prezintă în mod sistematizat în tabelul 14.8. Linia y_i şi respectiv coloana x_i reprezintă mijloacele intervalelor de variaţie ale celor două variabile. Linia 10 a tabelului 14.8 s-a determinat înmulţind fiecare valoare a lui y_i cu toate valorile lui x_i şi cu frecvenţele de la intersecţia lui y_i cu rândul lui x_i, adică cu n_{ij}, astfel:

$$\sum_{i=1}^{5} y_1 x_i n_{i1} = 25\cdot 5\cdot 3 + 25\cdot 15\cdot 0 + 25\cdot 25\cdot 0 + 25\cdot 35\cdot 0 + 25\cdot 45\cdot 0 = 375$$

$$\sum_{i=1}^{5} y_2 x_i n_{i2} = 75\cdot 5\cdot 3 + 75\cdot 15\cdot 8 + 75\cdot 25\cdot 4 = 176252$$

$$\sum_{i=1}^{5} y_3 x_i n_{i3} = 125\cdot 15\cdot 3 + 125\cdot 25\cdot 6 + 125\cdot 35\cdot 2 = 33125$$

$$\sum_{i=1}^{5} y_4 x_i n_{i4} = 175\cdot 25\cdot 2 + 175\cdot 35\cdot 1 + 175\cdot 45\cdot 3 = 385003$$

$$\sum_{i=1}^{5} y_5 x_i n_{i4} = 225\cdot 35\cdot 1 + 225\cdot 45\cdot 3 = 38250$$

$$\sum_{j=1}^{5}\sum_{i=1}^{5} y_i x_i n_{ij} = 375 + 17625 + 33125 + 38500 + 38250 = 127875$$

În mod analog s-au determinat elementele de pe coloana 11 a tabelului 14.8. Se observă că , indiferent de ordinea însumării, cele două sume duble sunt egale:

$$\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} = \sum_{j=1}^{p}\sum_{i=1}^{k} y_j x_i n_{ij} \qquad i = 1,2,\ldots,k \qquad j = 1,2,\ldots,p$$

sau

$$\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} = \sum_{i=1}^{k} x_i \sum_{j=1}^{p} y_j n_{ij}$$

Scanned with OKEN Scanner

$$\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} = \sum_{j=1}^{p} y_j \sum_{i=1}^{k} x_i n_{ij}$$

deci

$$\sum_{i=1}^{k} x_i \sum_{j=1}^{p} y_j n_{ij} = \sum_{j=1}^{p} y_j \sum_{i=1}^{k} x_i n_{ij} = \sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij}$$

În cazul ecuaţiei de regresie $y_x = a + bx_i$, parametrii a şi b se calculează pe baza relaţiilor (14.24) şi (14.25):

$$a = \frac{27175 \cdot 4525 - 905 \cdot 127875}{39 \cdot 27175 - 905^2} = \frac{289600}{240800} = 30{,}066$$

$$b = \frac{39 \cdot 127875 - 905 \cdot 4525}{39 \cdot 27175 - 905^2} = \frac{891999 \cdot 99}{240800} = 3{,}7043$$

Deci, legătura dintre distribuţia personalului pe judeţe şi distribuţia fondurilor fixe este descrisă de ecuaţia:

$$y_x = 30{,}066 + 3{,}7043\, x_i$$

Ecuaţia dreptei y_x nu este o ecuaţie generală a oricărei linii drepte, care descrie legătura dintre cele două variabile, ci reprezintă doar o ecuaţie particulară a dreptei faţă de care suma abaterilor valorilor empirice y_i faţă de cele teoretice este minimă. Dacă *abaterile în jurul acestei drepte sunt normal distribuite, dreapta va fi cea mai potrivită pentru regresie.* Dar, dreapta de regresie obţinută pe baza metodei celor mai mici pătrate nu este în mod necesar cea mai bună ecuaţie de regresie, deoarece:

1. S-ar putea ca legătura dintre variabile să nu fie liniară, caz în care ecuaţia liniară de regresie nu este cea mai potrivită ecuaţie de regresie, ci doar cea mai bună dintre ecuaţiile de formă liniară;

2. Mai mult decât atât, nu este în mod necesar cea mai bună dreaptă de regresie. Ea are doar o şansă mai bună de a fi mai potrivită decât oricare altă dreaptă;

3. Dacă abaterile $y_i - y_{x_i}$ nu sunt normal distribuite, atunci nu avem temei să considerăm că dreapta obţinută pe baza metodei celor mai mici pătrate este mai bună decât cele obţinute pe baza altor metode.

Scanned with OKEN Scanner

În practica statistică, metoda celor mai mici pătrate este cea mai des utilizată pentru estimarea parametrilor ecuaţiilor de regresie, deoarece este simplă, obiectivă şi indicatorii de regresie se pot interpreta probabilistic.

14.2. Dualitatea în regresie

În cazul a două variabile există două drepte de regresie y_x şi x_y, care, determinate cu ajutorul metodei celor mai mici pătrate, au coeficienţi unghiulari diferiţi, astfel:

- pentru dreapta de regresie $y_x = a + b_1 x_i$, coeficientul de regresie are expresia:

$$b_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (14.26)$$

- pentru dreapta de regresie $x_y = a + b_2 y_i$, coeficientul de regresie va fi:

$$b_2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (14.27)$$

În primul caz, abaterile valorilor empirice y_i faţă de cele corespunzătoare de pe dreapta de regresie y_{x_i} se consideră paralele cu axa ordonatelor, iar în cazul al doilea cu axa absciselor (fig.14.4).

Reprezentând pe acelaşi grafic cele două drepte de regresie se obţine figura14.4. Dreapta y_x se va situa , de regulă, sub x_y , deoarece coeficientul de regresie b_1 este de cele mai multe ori mai mic decât b_2. Cele două drepte de regresie, reprezentate pe acelaşi grafic, formează între ele un unghi de o anumită mărime. Când acest unghi este foarte mare regresia este foarte puternică şi legătura dintre variabile foarte slabă, dreptele de regresie apropiindu-se în mai mare măsură de axele de coordonate şi invers. În cazul

Scanned with OKEN Scanner

limită opus, când cele două drepte de regresie coincid, regresia este nulă, iar legătura dintre variabile este funcţională.

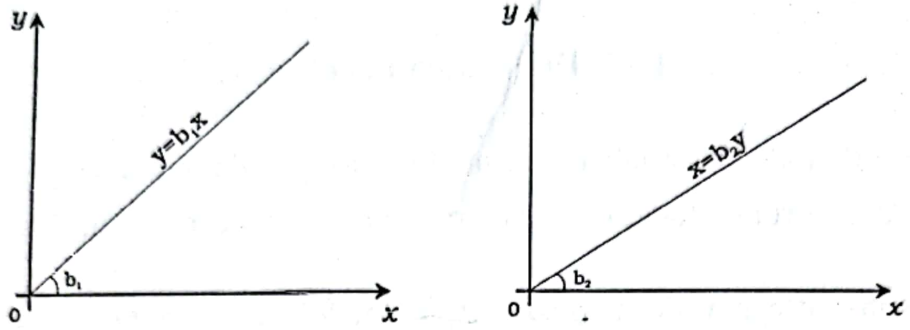


Figura 14.4. Dualitatea în regresie

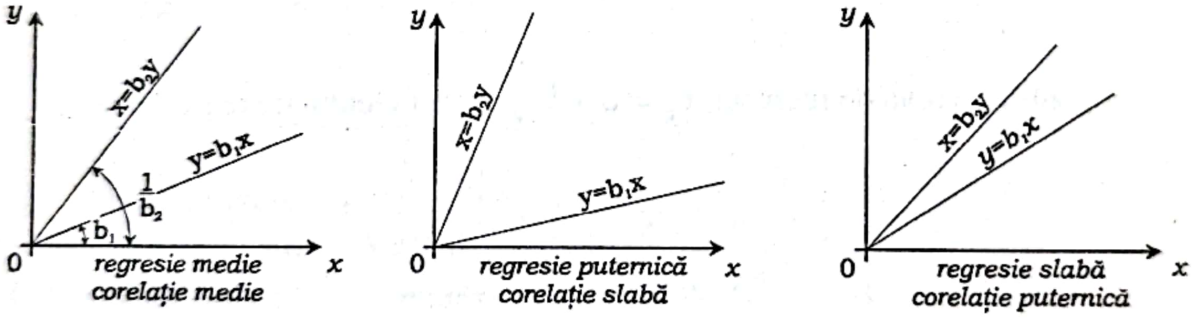


Figura 14.5. Corespondenţa dintre regresie şi corelaţie

Dacă dreapta de regresie trece prin origine (fig. 14.4 şi fig. 14.5) atunci ea depinde numai de un singur parametru şi anume, de coeficientul de regresie b şi va avea forma:

$$y_x = b_1 x \qquad \text{şi} \qquad x_y = b_2 y$$

Scanned with OKEN Scanner

În acest caz particular, coeficientul de regresie se determină tot cu ajutorul metodei celor mai mici pătrate. De exemplu, pentru determinarea lui b_1 se va proceda astfel:

$$S = \sum_{i=1}^{n} (y_i - y_{x_i})^2 = minim$$

Substituind pe y_x cu $b_1 x$, rezultă:

$$S = \sum_{i=1}^{n} (y_i - b_1 x_i)^2 = minim$$

Calculând derivata întâi a lui S în raport cu b_1 şi egalând-o cu zero, se obţine:

$$\frac{\partial S}{\partial b_1} = -2\sum_{i=1}^{n} (y_i - b_1 x_i) x_i = 0$$

şi

$$\sum_{i=1}^{n} y_i x_i = b_1 \sum_{i=1}^{n} x_i^2$$

de unde

$$b_1 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2} \qquad (14.28)$$

În mod analog se determină coeficientul de regresie b_2 şi se obţine:

$$b_2 = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} y_i^2} \qquad (14.29)$$

Scanned with OKEN Scanner

Coeficientul b_2 se calculează în practică numai în cazul în care se poate considera că între variabilele x şi y există o interacţiune reciprocă reală.

În cazul mai general al regresiei liniare, când $a \neq 0$, parametrii dreptei de regresie $x_y = a_2 + b_2 y$ se determină tot cu ajutorul metodei celor mai mici pătrate, obţinându-se sistemul de ecuaţii normale:

$$\begin{aligned} na_2 + b_2 \sum_{i=1}^{n} y_i &= \sum_{i=1}^{n} x_i \\ a_2 \sum_{i=1}^{n} y_i + b_2 \sum_{i=1}^{n} y_i^2 &= \sum_{i=1}^{n} y_i x_i \end{aligned} \qquad (14.30)$$

de unde prin rezolvare rezultă:

$$a_2 = \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i^2 - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i x_i}{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}$$

$$b_2 = \frac{n \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n \sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2}$$

Să considerăm, pentru exemplificarea dualităţii în regresie, legătura dintre cantitatea de cărbune extras şi producţia de energie termoelectrică (tabelul 14.1).

Parametrii ecuaţiei de regresie a variabilei X, în funcţie de variabila Y, se determină cu ajutorul formulelor (14.23) şi (14.24):

$$a_2 = \frac{309{,}2 \cdot 14558{,}08 - 442{,}6 \cdot 9726{,}46}{17 \cdot 14558{,}08 - 442{,}6^2} = 3{,}807$$

$$b_2 = \frac{17 \cdot 9726{,}46 - 309{,}2 \cdot 442{,}6}{17 \cdot 14558{,}08 - 442{,}6^2} = 0{,}552$$

Scanned with OKEN Scanner

a_2 = 3,807 exprimă faptul că dreapta de regresie x_y intersectează axa ox în punctul de coordonate $y = 0$ şi x = 3,807 milioane tone.

b_2 = 0,552 milioane tone, înseamnă că producţia de cărbune creşte cu peste jumătate de milion de tone, atunci când producţia de energie termoelectrică creşte cu un miliard de kwh. Ecuaţia de regresie va fi:

$$x_y = 3{,}807 + 0{,}552\, y_i$$

Valorile teoretice ale lui x_y corespunzătoare diferitelor valori ale variabilei Y, s-au înscris în coloana 7 a tabelului 14.3.

14.3. Intensitatea legăturilor liniare

14.3.1. Coeficientul de corelaţie

Determinarea ecuaţiei liniare de regresie, în fiecare caz particular, oferă imaginea formei legăturii dintre două variabile, dar nu răspunde la întrebarea, foarte importantă şi anume cât de puternică sau cât de intensă este legătura sau dependenţa dintre variabilele corelate. Pentru a rezolva această problemă, să considerăm o distribuţie bidimensională situată în cadranul întâi al axelor de coordonate rectangulare (fig. 14.6).

Dreptele $x = \bar{x}$ şi $y = \bar{y}$ corespunzătoare mediilor variabilelor X şi Y împart cadranul întâi în patru sectoare, în care abaterile variantelor individuale ale celor două variabile faţă de mediile lor pot fi pozitive sau negative, astfel:

- în sectorul I, abaterile variantelor variabilei X faţă de media lor $\bar{x}$ sunt negative, iar cele corespunzătoare variabilei Y sunt pozitive:

$$(x_i - \bar{x}) < 0 \qquad \text{şi} \qquad (y_i - \bar{y}) > 0$$

- în sectorul II, abaterile ambelor variabile faţă de mediile lor sunt pozitive:

$$(x_i - \bar{x}) > 0 \qquad \text{şi} \qquad (y_i - \bar{y}) > 0$$

- în sectorul III, abaterile variabilei X sunt pozitive, iar cele ale variabilei Y sunt negative:

Scanned with OKEN Scanner

$$(x_i - \bar{x}) > 0 \qquad \text{şi} \qquad (y_i - \bar{y}) < 0$$

- în sectorul *IV*, abaterile ambelor variabile sunt negative:

$$(x_i - \bar{x}) < 0 \qquad \text{şi} \qquad (y_i - \bar{y}) < 0$$

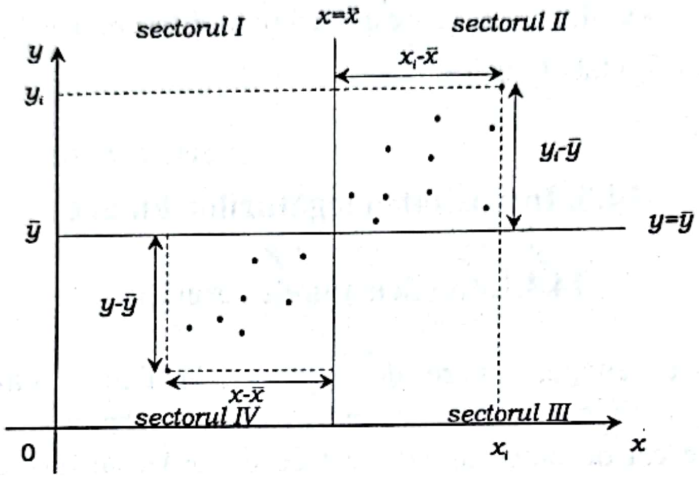


Figura 14.6. Mediile şi abaterile a două variabile corelate

Considerând acum produsele abaterilor celor două variabile şi suma lor pentru fiecare sector, observăm că acestea sunt pozitive în sectoarele *II* şi *IV* şi negative în sectoarele *I* şi *III*.

Când majoritatea sau chiar totalitatea punctelor corespunzătoare perechilor de valori corelate x_i, y_i se situează în sectoarele *II* şi *IV*, atunci înseamnă că între variabilele *X* şi *Y* există o corelaţie liniară directă şi puternică.

Când majoritatea sau chiar totalitatea punctelor se situează în sectoarele *I* şi *III*, corelaţia liniară dintre variabile va fi puternică şi inversă (sau indirectă).

În sfârşit, dacă o parte din puncte se situează în sectoarele *II* şi *IV* şi o altă parte în *I* şi *III* aceasta sugerează existenţa unei distribuţii întâmplătoare a punctelor în cadranul *I* al axelor de coordonate, ceea ce înseamnă o legătură slabă între variabile.

Intensitatea legăturii liniare dintre două variabile se poate măsura cu ajutorul *coeficientului de corelaţie simplă liniară* - notat cu *r* care se poate calcula ca o medie aritmetică simplă a produsului abaterilor normale normate ale celor două variabile.

Scanned with OKEN Scanner

Dacă notăm abaterile normale normate:

$$z_x = \frac{x_i - \bar{x}}{\sigma_x} \qquad \text{şi} \qquad z_y = \frac{y_i - \bar{y}}{\sigma_y}$$

atunci $-1 \le r_{xy} \le 1$

$$r_{xy} = r_{yx} = \overline{z_x \cdot z_y} = \frac{\sum_{i=1}^{n} \frac{x_i - \bar{x}}{\sigma_x} \cdot \frac{y_i - \bar{y}}{\sigma_y}}{n} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x \sigma_y} \qquad (14.31)$$

Coeficientul de corelaţie este o mărime abstractă, independentă de unităţile de măsură ale celor două variabile, fapt care-i măreşte posibilităţile de comparabilitate.

Când $r = +1$, ecuaţiile de regresie y_x şi x_y coincid, ceea ce înseamnă că legătura dintre variabile este de tip funcţional, adică fiecărei valori date ale lui X îi corespunde o valoare pentru Y şi numai una. În acest caz, ambele variabile variază în acelaşi sens, iar legătura dintre ele poartă denumirea de *corelaţie pozitivă* (*directă*) *perfectă.*

În cazul opus, când $r = -1$, perechile de valori corelate x_i, y_i se înscriu tot pe o dreaptă, ca şi în cazul anterior, cu deosebirea că valorile mari ale lui X corespund valorilor mici ale lui Y şi invers. Legătura dintre variabile este tot de tip funcţional şi poartă denumirea de *corelaţie negativă* (*inversă*) *perfectă.*

Dacă variabilele studiate sunt independente, $r = 0$ deoarece, după cum vom vedea în continuare, coeficienţii de regresie ai celor două drepte $y_x = a_1 + b_1 x$ şi $x_y = a_2 + b_2 y$ sunt egali cu zero, ceea ce înseamnă că dreptele de regresie sunt paralele cu axele de coordonate. Reciproca nu este întotdeauna valabilă. Dacă $r = 0$, nu înseamnă că cele două variabile sunt independente în general, ci sunt numai necorelate liniar. Iată un exemplu care sugerează mai bine această idee. Să considerăm că perechile de valori x_i, y_i se prezintă ca în figura 14.7.

Scanned with OKEN Scanner

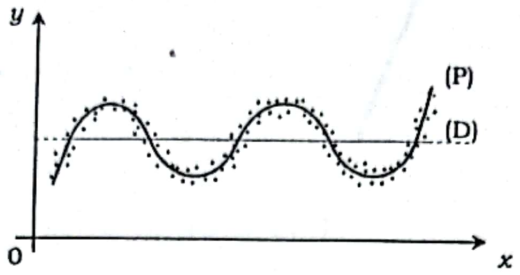


Figura 14.7. Corelaţie curbilinie puternică şi lipsă a unei legături liniare

În acest caz, cea mai bună dreaptă de regresie este dreapta D paralelă cu axa ox, deci cu coeficientul de regresie nul şi $r = 0$. Dacă se renunţă la modelul liniar de regresie şi se consideră curba sinusoidală P, se observă că aceasta reprezintă o legătură sinusoidală sau ciclică perfectă. Ar fi, deci, greşit să considerăm că, dacă $r = 0$, nu există nici un fel de legătură între variabile.

Cu cât r se apropie mai mult de cele două valori extreme ± 1, cu atât intensitatea legăturii este mai puternică şi invers. Dacă $r = 0$, aceasta înseamnă că între variabile *nu există o legătură liniară*, dar poate să existe o legătură curbilinie. La interpretarea coeficientului de corelaţie, ca de altfel a oricărui indicator al intensităţii corelaţiei, trebuie să se ţină seama de numărul perechilor de valori corelate.

Pentru o mai mare stabilitate şi pentru a nu depinde prea mult de fluctuaţiile de selecţie, indicatorii de regresie şi corelaţie trebuie să se calculeze pentru un număr suficient de mare de valori ale variabilelor care se corelează. Aceasta înseamnă că numărul variantelor variabilelor trebuie să fie mai mare sau cel puţin egal cu 30. Astfel, dacă numărul de valori ale variabilelor luate în calcul este foarte mare, un coeficient de corelaţie foarte mic să zicem 0,3 sau 0,4 va arăta desigur o legătură slabă, dar stabilă, reală. Dacă nu dispunem decât de un număr mic de perechi de valori, să zicem 10, atunci un coeficient de corelaţie ce se apropie de unitate nu va avea decât o semnificaţie îndoielnică.

Formula (14.31) de calcul a coeficientului de corelaţie se poate utiliza în practică, atunci când variantele celor două variabile X şi Y se prezintă sub forma a două serii paralele. Determinarea acestui indicator poate fi efectuată mai uşor pe baza unei formule de calcul simplificat.

Pentru determinarea acestei formule, se poate porni de la relaţia (14.31) substituind pe $\bar{x}$, $\bar{y}$, σ_x şi σ_y 200 cu expresiile lor dezvoltate şi apoi efectuând calculele se obţine:

$$r_{x_i,y_i} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{n\sigma_x\sigma_y} = \frac{\sum_{i=1}^{n}(x_iy_i - x_i\bar{y} - \bar{x}y_i + \bar{x}\bar{y})}{n\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}\cdot\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}}} =$$

$$= \frac{\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\bar{y} - \sum_{i=1}^{n}\bar{x}y_i - \sum_{i=1}^{n}\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}x_i\bar{x} + \sum_{i=1}^{n}\bar{x}^2)(\sum_{i=1}^{n}y_i^2 - 2\sum_{i=1}^{n}y_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2)}} =$$

$$= \frac{\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\frac{\sum_{i=1}^{n}y_i}{n} - \frac{\sum_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}y_i + n\frac{\sum_{i=1}^{n}x_i}{n}\frac{\sum_{i=1}^{n}y_i}{n}}{\sqrt{(\sum_{i=1}^{n}x_i^2 - 2\sum_{i=1}^{n}x_i\frac{\sum_{i=1}^{n}x_i}{n} + n\cdot\frac{\sum_{i=1}^{n}x_i}{n}\cdot\frac{\sum_{i=1}^{n}x_i}{n})(\sum_{i=1}^{n}y_i^2 - 2\frac{\sum_{i=1}^{n}y_i}{n}\cdot\sum_{i=1}^{n}y_i + n\cdot\frac{\sum_{i=1}^{n}y_i}{n}\cdot\frac{\sum_{i=1}^{n}y_i}{n})}}$$

de unde rezultă:

$$r_{x_i,y_i} = \frac{n\sum_{i=1}^{n}x_iy_i - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{\left[n\sum_{i=1}^{n}x_i^2 - (\sum_{i=1}^{n}x_i)^2\right]\cdot\left[n\sum_{i=1}^{n}y_i^2 - (\sum_{i=1}^{n}y_i)^2\right]}} \qquad (14.32)$$

Calculul coeficientului de corelaţie simplă se simplifică în mare măsură, dacă se utilizează formula (14.32), deoarece toate sumele ce intervin în formulă sunt deja determinate cu ocazia calculului parametrilor ecuaţiei de regresie (tabelul 14.3).

Să considerăm, pentru exemplificarea calculului coeficientului de corelaţie liniară, exemplul din tabelul 14.1, referitor la legătura dintre producţia de energie termoelectrică şi producţia de cărbune extras în *România* în perioada 1960-1976.

Schematic, calculele intermediare pentru determinarea coeficientului de corelaţie se prezintă în tabelul 14.9.

Scanned with OKEN Scanner

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{309,2}{17} = 18,188 \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{442,6}{17} = 26,035$$

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}} = \sqrt{\frac{960,318}{17}} = 7,5159 \text{ } \textit{milioane tone cărbune}$$

Tabelul 14.9. Calculele intermediare necesare pentru determinarea lui r

Anul	Mil.t. x_i	mld. kwh y_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
0	1	2	3	4	5	6	7
1960	8,2	7,3	-9,988	99,760	-18,735	351,000	187,125
1961	8,7	8,2	-9,488	90,022	-17,835	318,087	169,218
1962	9,6	9,4	-8,588	73,754	-16,635	276,723	153,167
1963	10,3	11,1	-7,888	62,221	-14,935	223,054	117,807
1964	11,1	13,3	-7,088	50,240	-12,735	162,180	90,266
1965	12,1	16,2	-6,088	37,064	-9,835	96,727	59,875
1966	13,5	19,8	-4,688	21,977	-6,235	38,875	29,230
1967	15,0	23,3	-3,188	10,163	-2,735	7,480	8,719
1968	17,0	26,3	-1,188	1,411	0,265	0,070	0,315
1969	19,2	29,3	1,012	1,024	3,265	10,660	3,304
1970	22,8	32,3	4,612	21,271	6,265	39,250	28,894
1971	23,0	35,0	4,812	23,155	8,965	80,371	43,140
1972	25,3	36,1	7,112	50,581	10,065	101,304	71,582
1973	26,7	39,2	8,512	72,454	13,165	173,317	112,060
1974	29,2	40,6	11,012	121,264	14,565	212,139	160,390
1975	29,4	45,0	11,212	125,709	18,965	359,671	212,636
1976	28,1	50,2	9,912	98,248	24,165	583,947	239,523
Total	309,2	442,6	-	960,318	-	3034,855	1687,251

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n}} = \sqrt{\frac{3034,855}{17}} = 13,361 \text{ } \textit{milioane tone cărbune}$$

Coeficientul de corelaţie liniară se determină cu ajutorul formulei 14.31:

Scanned with OKEN Scanner

$$r_{x_i,y_i} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y} = \frac{1687,251}{17 \cdot 7,5159 \cdot 13,361} = 0,98$$

$r_{x_i,y_i} = 0,98$ arată că între producţia de cărbune şi producţia de energie termoelectrică există o corelaţie directă foarte intensă.

Calculul coeficientului de corelaţie simplă liniară este mult simplificat, dacă se utilizează formula (14.32), deoarece toate sumele ce intervin în formulă sunt deja determinate cu ocazia calculului parametrilor ecuaţiei de regresie (tabelul 14.3).

$$r_{x_i,y_i} = \frac{17 \cdot 9726,46 - 309,2 \cdot 442,6}{\sqrt{(17 \cdot 6584,12 - 309,2^2)(17 \cdot 14558,08 - 442,6^2)}} = \frac{28497,9}{29021,877} = 0,98$$

Se observă că s-a obţinut acelaşi rezultat, ca şi în cazul utilizării formulei 14.31, deoarece formulele 14.31 şi 14.32 sunt echivalente.

Dacă perechile de valori x_i, y_i ale celor două variabile empirice sunt ponderate cu frecvenţa n_i, atunci formula coeficientului de corelaţie (14.31) devine:

$$r_{x_i,y_i} = \frac{\sum_{i=1}^{k}(x_i - \bar{x})(y_i - \bar{y})n_i}{(\sum_{i=1}^{k} n_i)\sigma_x \sigma_y} \qquad (14.33)$$

în care:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_i}{\sum_{i=1}^{k} n_i}} \quad \text{şi} \quad \sigma_x = \sqrt{\frac{\sum_{i=1}^{k}(y_i - \bar{y})^2 n_i}{\sum_{i=1}^{k} n_i}}$$

Similar cu cazul anterior, se va obţine o formulă de calcul simplificat al coeficientului de corelaţie, astfel:

$$r_{x_i,y_i,n_i} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} x_i y_i n_i - \sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} y_i n_i}{\sqrt{\left[\sum_{i=}^{k} n_i \sum_{i=1}^{k} x_i^2 n_i - (\sum_{i=1}^{k} x_i n_i)^2\right] \cdot \left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} y_i^2 n_i - (\sum_{i=1}^{k} y_i n_i)^2\right]}} \quad (14.34)$$

În cazul prezentării datelor sub forma unui tabel de corelaţie, formula coeficientului de corelaţie (14.31) devine:

$$r_{x_i,y_j,n_{ij}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{s}(x_i - \bar{x})(y_j - \bar{y})n_{ij}}{(\sum_{i=1}^{k}\sum_{j=1}^{s} n_{ij})\sigma_x \sigma_y} \quad (14.35)$$

în care:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2 n_{i.}}{\sum_{i=1}^{k} n_{i.}}} \quad \text{şi} \quad \sigma_y = \sqrt{\frac{\sum_{j=1}^{s}(y_j - \bar{y})^2 n_{.j}}{\sum_{j=1}^{p} n_{.j}}}$$

Folosirea formulei (14.35) pentru calculul coeficientului de corelaţie pe baza datelor din tabelul 14.8 necesită întocmirea în prealabil a tabelului 14.11.

Pe baza datelor din tabelul 14.8 se pot determina mediile celor două variabile:

$$\bar{x} = \frac{\sum_{i=1}^{k} x_i n_{i.}}{\sum_{i=1}^{k} n_{i.}} = \frac{905}{39} = 23{,}205 \quad \text{şi} \quad \bar{y} = \frac{\sum_{j=1}^{p} y_j n_{j.}}{\sum_{j=1}^{p} n_{.j}} = \frac{4525}{39} = 116{,}026$$

Substituind datele din tabelul 14.10 în formula (14.35), rezultă:

$$r_{x_i,y_j,n_{ij}} = \frac{22871{,}794}{39 \cdot 12{,}582 \cdot 55{,}322} = 0{,}84$$

Coeficientul de corelaţie $r_{x_i,y_j,n_{ij}} = 0{,}84$ evidenţiază faptul că între numărul mediu al personalului şi valoarea fondurilor fixe există o corelaţie directă şi intensă.

Scanned with OKEN Scanner

Tabelul 14.11. Calculul parametrilor ecuaţiei regresiei simple liniare în cazul unei distribuţii bidimensionale

X \ Y	- 50	50 - 100	100 - 150	150 - 200	200 -	Total ($n_{i.}$)	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 n_{i.}$	$\sum_{j=1}^{s}(x_i - \bar{x})(y_j - \bar{y})n_{ij}$
1	2	3	4	5	6	7	8	9	10	11	12
- 10	3	3	-	-	-	6	5	-18,205	331,422	1988,532	7212,020
10 - 20	-	8	3	-	-	11	15	-8,205	67,322	740,542	2472,052
20 - 30	-	4	6	2	-	12	25	1,795	3,222	38,664	13,800
30 - 40	-	-	2	1	1	4	35	11,795	139,122	556,488	2192,643
40 -	-	-	-	3	3	6	45	21,795	475,022	2850,132	10981,279
Total ($n_{.j}$)	3	15	11	6	4	39	-			6174,358	22871,794
y_j	25	75	125	175	225	-					
$y_j - \bar{y}$	-91,026	-41,0260	8,974	58,974	108,974	-					
$(y_j - \bar{y})^2$	8285,733	1683,1326	80,533	3477,933	11875,332	-					
$(y_j - \bar{y})^2 n_{.j}$	24857,197	25246,989	885,859	20867,595	47501,328	1119358,968					
$\sum_{i=1}^{k}(x_i - \bar{x})(y_j - \bar{y})n_{ij}$	4971,385	4639,015	87,452	4763,330	8410,612	22871,794					

Scanned with OKEN Scanner

Pe baza formulei (14.35), prin transformări elementare, se obţine formula (14.36) de calcul simplificat a coeficientului de corelaţie:

$$r_{x_i,y_j,n_{ij}} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} - \sum_{i=1}^{k} x_i n_{i.} \sum_{j=1}^{p} y_j n_{.j}}{\sqrt{\left[\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{i=1}^{k} x_i^2 n_{i.} - (\sum_{i=1}^{k} x_i n_{i.})^2\right]\left[\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{j=1}^{p} y_j^2 n_{.j} - (\sum_{j=1}^{p} y_j n_{.j})^2\right]}} \tag{14.36}$$

O simplificare mai mare a calculelor se va obţine, dacă se substituie valorile x_i şi y_i ale variabilelor cu x_i', respectiv y_i', date de relaţiile:

$$x_i' = \frac{x_i - \bar{x}}{k_x} \qquad \text{şi} \qquad y_i' = \frac{y_i - \bar{y}}{k_y} \tag{14.37}$$

unde x_0, respectiv y_0, reprezintă nivelurile variabilelor din centrul celor două distribuţii, iar k_x şi k_y reprezintă cel mai mare divizor comun al diferenţelor faţă de originea de calcul aleasă.

Substituind pe (14.37) în relaţiile (14.32), (14.34) şi (14.36) se pot obţine alte formule de calcul simplificat ale coeficientului de corelaţie:

- în cazul când perechile de valori x_i, y_i sunt neponderate:

$$r_{x_i,y_i} = \frac{n\sum_{i=1}^{n} \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} - \sum_{i=1}^{n} \frac{x_i - x_0}{k_x} \cdot \sum_{i=1}^{n} \frac{y_i - y_0}{k_y}}{\sqrt{\left[n\sum_{i=1}^{n}\left(\frac{x_i - x_0}{k_x}\right)^2 - \left(\sum_{i=1}^{n} \frac{x_i - x_0}{k_x}\right)^2\right]\left[n\sum_{i=1}^{n}\left(\frac{y_i - y_0}{k_y}\right)^2 - \left(\sum_{i=1}^{n} \frac{y_i - y_0}{k_y}\right)^2\right]}} \tag{14.38}$$

- în cazul ponderării perechilor de valori corelate x_i, y_i cu n_i :

$$r_{x_i y_i n_i} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \frac{x_i - x_0}{k_x} \cdot \frac{y_i - y_0}{k_y} n_i - \sum_{i=1}^{k} \frac{x_i - x_0}{k_x} n_i \sum_{i=1}^{k} \frac{y_i - y_0}{k_y} n_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \left(\frac{x_i - x_0}{k_x}\right)^2 n_i - \left(\sum_{i=1}^{k} \frac{x_i - x_0}{k_x} n_i\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \left(\frac{y_i - y_0}{k_y}\right)^2 n_i - \left(\sum_{i=1}^{k} \frac{y_i - y_0}{k_y} n_i\right)^2\right]}} \tag{14.39}$$

• în cazul unui tabel de corelaţie:

$$r = \frac{\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{i=1}^{k}\sum_{j=1}^{p} \frac{x_i - x_0}{k_x} \cdot \frac{y_j - y_0}{k_y} n_{ij} - \sum_{i=1}^{k} \frac{x_i - x_0}{k_x} n_{i.} \cdot \sum_{j=1}^{p} \frac{y_j - y_0}{k_y} n_{.j}}{\sqrt{\left[\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{i=1}^{k} \left(\frac{x_i - x_0}{k_x}\right)^2 n_{i.} - \left(\sum_{i=1}^{k} \frac{x_i - x_0}{k_x} n_{i.}\right)^2\right]\left[\sum_{i=1}^{k}\sum_{j=1}^{p} n_{ij} \sum_{j=1}^{p} \left(\frac{y_j - y_0}{k_y}\right)^2 n_{.j} - \left(\sum_{j=1}^{p} \frac{y_j - y_0}{k_y} n_{.j}\right)^2\right]}} \tag{14.40}$$

Să exemplificăm utilizarea formulei (14.40) pe baza datelor din tabelul 14.7. Sistematizarea calculelor preliminare necesare determinării coeficientului de corelaţie în acest caz se prezintă în tabelul 14.11.

Considerând: $x_0 = x_3 = 25$ şi $y_0 = y_3 = 125$, iar ca divizori comuni ai diferenţelor $x_i - x_0$ şi $y_i - y_0$ vor fi $k_x = 10$, iar $k_y = 50$.

Substituind elementele de calcul din tabelul 14.11 în formula (14.40), se obţine:

$$r = \frac{39 \cdot 47 - (-7)(-7)}{\sqrt{\left[39 \cdot 63 - (-7)^2\right]\left[39 \cdot 49 - (-7)^2\right]}} = \frac{1784}{2117{,}4739} = 0{,}84$$

S-a obţinut aceeaşi valoare a coeficientului de corelaţie ca şi în cazul aplicării formulei (14.35), dar cu un volum de muncă mult mai redus. Astfel, pentru efectuarea calculelor din tabelul 14.11 nu este necesar să se apeleze la nici un mijloc modern de calcul, deoarece se lucrează cu numere mici şi întregi, iar calculele se pot face mintal.

Scanned with OKEN Scanner

Tabelul 14.12. Calculul simplificat al lui r în cazul unei distribuţii bidimensionale

	X \ Y	-50	50 - 100	100 -150	150 -200	200 -	$n_{i.}$	x_i	$x_i - x_0$	$\frac{x_i - x_0}{k_y}$	$\frac{x_i - x_0}{k_y} n_i$	$\frac{(x_i - x_0)^2}{k_y} n_i$	$\sum_{j=1}^{s} \frac{x_i - x_0}{k_x} \cdot \frac{y_j - y_0}{k_y}$
0	1	2	3	4	5	6	7	8	9	10	11	12	13
1	-10	3	3	-	-	-	6	5	-20	-2	-12	24	18
2	10 - 20	-	8	3	-	-	11	15	-10	-1	-11	11	8
3	20 - 30	-	4	6	2	-	12	25	0	0	0	0	0
4	30 - 40	-	-	2	1	1	4	35	10	1	4	4	3
5	40	-	-	-	3	3	6	45	20	2	12	24	18
6	$n_{.j}$	3	15	11	6	4	39	-	-	-	-7	63	47
7	y_j	25	75	125	175	225	-						
8	$y_j - \bar{y}$	-100	-50	0	50	100	-						
9	$\frac{y_j - y_0}{k_y}$	-2	-1	0	1	2	-						
10	$\frac{y_j - y_0}{k_y} n_{.j}$ 209	-6	-15	0	6	8	-7						
11	$\frac{(y_j - y_0)^2}{k_y} n_{.j}$	12	15	0	6	16	49						
12	$\sum_{i=1}^{k} \frac{x_i - x_0}{k_x} \cdot \frac{y_j - y_0}{k_y} n_{ij}$	12	14	0	7	14	47						

Scanned with OKEN Scanner

14.3.2. Raportul de corelaţie

Deoarece coeficientul de corelaţie este un indicator al intensităţii legăturii numai în cazul celor liniare, cunoscând valorile empirice ale variabilei rezultative y_i şi valorile teoretice corespunzătoare y_x se poate stabili un indicator general valabil al intensităţii legăturii dintre fenomene, numit *raport de corelaţie*, pe care îl vom nota în continuare cu η. Calculul lui η se bazează pe descompunerea dispersiei totale a variabilei empirice Y pe factori de influenţă.

Variaţia variabilei empirice Y depinde de mărimea influenţei tuturor factorilor cauzali care acţionează asupra ei, inclusiv cea a factorului X, în timp ce variaţia variabilei teoretice Y_x depinde numai de influenţa factorului X. Aceasta înseamnă că *dispersia totală a variabilei empirice Y se poate descompune în dispersia valorilor empirice faţă de valorile ei teoretice şi dispersia valorilor teoretice Y_x* faţă de media lor.

Abaterile valorilor empirice faţă de media lor se pot descompune în două componente şi anume abaterile valorilor empirice faţă de valorile teoretice şi abaterile valorilor teoretice faţă de media lor:

$$y_i - \bar{y} = (y_i - y_x) + (y_x - \bar{y})$$

Sugestiv, aceste abateri se prezintă în figura 14.8.

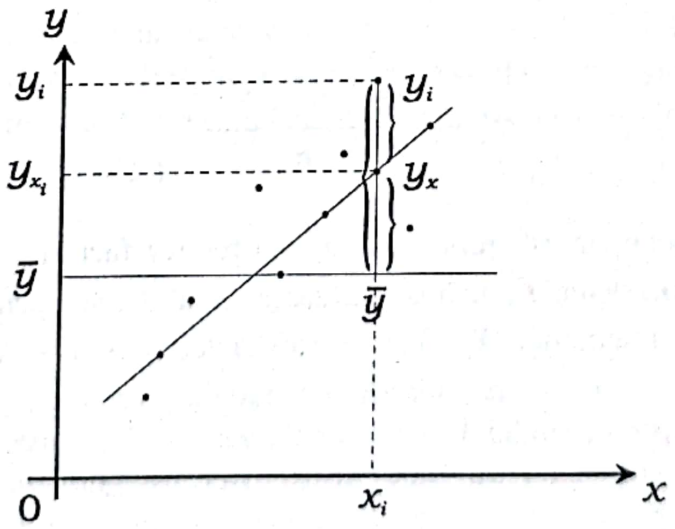


Figura 14.8. Descompunerea variaţiei totale în componentele ei

Scanned with OKEN Scanner

Ridicând la pătrat abaterile şi însumând, rezultă:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-y_x)^2+\sum_{i=1}^{n}(y_x-\bar{y})^2$$

Împărţind ambii membri ai acestei relaţii prin numărul perechilor de valori corelate n, rezultă următoarele dispersii:

- *dispersia generală* a variabilei Y:

$$\sigma_y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}$$

- dispersia valorilor teoretice ale variabilei faţă de media lor, numită *dispersie explicată*:

$$\sigma_{y_x}^2=\frac{\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2}{n}$$

- dispersia valorilor empirice faţă de valorile teoretice ale variabilei Y, numită şi *dispersie reziduală*:

$$\sigma_{yy_x}^2=\frac{\sum_{i=1}^{n}(y_i-y_{x_i})^2}{n}$$

Dispersia generală σ_y^2 reflectă influenţa tuturor factorilor care au acţionat asupra variabilei dependente Y, inclusiv influenţa variabilei independente X, în timp ce dispersia valorilor teoretice Y_{x_i} faţă de media lor $\bar{y}$ măsoară numai influenţa variabilei independente X asupra variabilei dependente, iar influenţa celorlalţi factori care au acţionat asupra variabilei Y, cu excepţia variabilei X, este dată de dispersia reziduală. Aceasta înseamnă că între aceste trei categorii de dispersii există următoarea relaţie(fig.14.9):

$$\sigma_y^2=\sigma_{y_x}^2+\sigma_{yy_x}^2 \qquad (14.41)$$

Scanned with OKEN Scanner

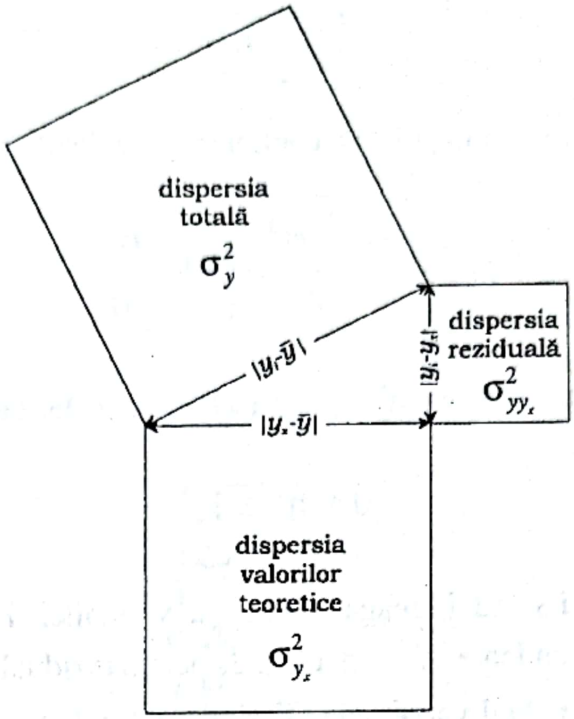


Figura 14.9. Relaţia dintre dispersiile unei variabile

În cazul unor legături funcţionale între variabilele X şi Y întreaga variaţie a variabilei dependente se datorează influenţei variabilei independente X şi, prin urmare, dispersia reziduală este egală cu zero. În această situaţie, valorile empirice y_i coincid cu valorile teoretice y_x:

$$y_i = y_x \qquad \text{şi} \qquad \sigma_y^2 = \sigma_{y_x}^2 \qquad \text{iar} \qquad \sigma_{yy_x}^2 = 0$$

Gradul de dependenţă dintre variabilele X şi Y sau intensitatea legăturii dintre ele se poate aprecia pe baza raportului dintre dispersia valorilor teoretice şi dispersia valorilor empirice ale variabilei Y.

Acest raport măsoară, deci, ponderea influenţei factorului specificat X asupra variaţiei variabilei Y, faţă de variaţia lui Y, pe seama tuturor factorilor de influenţă , inclusiv X. Indicatorul, astfel obţinut, poartă denumirea de *coeficient de determinaţie*, se notează cu η_d^2 şi se calculează pe baza relaţiei:

Scanned with OKEN Scanner

$$\eta_d^2 = \frac{\sigma_{y_x}^2}{\sigma_y^2} \tag{14.42}$$

ţinând seama de relaţia (14.41), coeficientul de determinaţie(14.42) devine:

$$\eta_d^2 = \frac{\sigma_y^2 - \sigma_{yy_x}^2}{\sigma_y^2} = 1 - \frac{\sigma_{yy_x}^2}{\sigma_y^2} \tag{14.43}$$

Din relaţia (14.30), rezultă că, acest coeficient de determinaţie η_d^2 variază între zero şi 1:

$$0 \le \eta_d^2 \le 1$$

$\eta_d^2 = 1$, atunci când întreaga variaţie a variabilei Y este determinată de variaţia variabilei independente X, deci când dispersia reziduală este egală cu zero.

$\eta_d^2 = 0$, atunci când variabilele X şi Y sunt independente şi deci dispersia reziduală este egală cu dispersia generală a variabilei Y.

Raportând dispersia reziduală la dispersia totală a variabilei Y, se obţine *coeficientul de nedeterminaţie* η_n^2:

$$\eta_n^2 = \frac{\sigma_{yy_x}^2}{\sigma_y^2} \tag{14.44}$$

şi

$$\eta_n^2 = \frac{\sigma_y^2 - \sigma_{y_x}^2}{\sigma_y^2} = 1 - \frac{\sigma_{y_x}^2}{\sigma_y^2} = 1 - \eta_d^2$$

deci

$$\eta_d^2 + \eta_n^2 = 1$$

Indicatorul intensităţii corelaţiei de acelaşi grad cu valorile empirice se obţine prin extragerea rădăcinii pătrate din coeficientul de determinaţie şi poartă denumirea de *raport de corelaţie*:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{\sigma_{y_x}^2}{\sigma_y^2}} \tag{14.45}$$

sau

$$\eta = \sqrt{1 - \frac{\sigma_{yy_x}^2}{\sigma_y^2}}$$

Substituind în formula (14.44) pe $\sigma_{y_x}^2$ şi σ_y^2 cu expresiile lor se obţine:

$$\eta = \sqrt{\frac{\sum_{i=1}^{n}(y_{x_i} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{14.46}$$

sau

$$\eta = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_{x_i})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{14.47}$$

Din formula (14.45) rezultă că raportul de corelaţie variază în intervalul [0,1]:

$$0 \le \eta \le 1$$

Calculul şi analiza raportului de corelaţie se vor efectua pe acelaşi exemplu pentru care s-a calculat coeficientul de corelaţie referitor la legătura dintre cantitatea de cărbune extras şi producţia de energie electrică (tabelul 14.12).

În paragraful 14.1.1 s-a obţinut ecuaţia de regresie a celor două variabile:

$$y_{x_i} = -5{,}7144 + 1{,}7456x_i$$

Scanned with OKEN Scanner

iar valorile teoretice ale variabilei Y s-au determinat în coloana 6 din tabelul 14.3.

Calculele preliminare necesare determinării raportului de corelaţie simplă liniară se prezintă, în mod sistematizat, în tabelul 14.12.

Substituind în formula (14.47) elementele de calcul din tabelul 14.12, rezultă:

$$\eta = \sqrt{1 - \frac{108,392}{3034,855}} = \sqrt{0,9643} = 0,98$$

Se observă că raportul de corelaţie $\eta = 0,98$ are aceeaşi valoare ca şi coeficientul de corelaţie calculat pe baza aceloraşi date. Coeficientul de determinaţie $\eta_d^2 = 96,43\%$ evidenţiază faptul că $96,43\%$ din variaţia producţiei de energie termoelectrică se datorează variaţiei producţiei de cărbune şi numai $3,57\%$ altor factori.

Tabelul 14.12. Elementele de calcul preliminar ale lui η

Anul	x_i	y_i	y_x	$(y_i - y_x)^2$	$(y_i - \bar{y})^2$
0	1	2	3	4	5
1960	8,2	7,3	8,599	1,687	351,000
1961	8,7	8,2	9,972	1,618	318,087
1962	9,6	9,4	11,043	2,699	276,723
1963	10,3	11,1	12,265	1,357	223,054
1964	11,1	13,3	13,662	0,131	162,180
1965	12,1	16,2	15,407	0,629	96,727
1966	13,5	19,8	17,851	3,799	38,875
1967	15,0	23,3	20,470	8,009	7,480
1968	17,0	26,3	23,961	5,471	0,070
1969	19,2	29,3	27,801	2,247	10,660
1970	22,8	32,3	34,085	3,186	39,250
1971	23,0	35,0	34,434	0,320	80,371
1972	25,3	36,1	38,449	5,518	101,304
1973	26,7	39,2	40,893	2,866	173,317
1974	29,2	40,6	45,257	21,688	212,139
1975	29,4	45,0	45,606	0,066	359,671
1976	28,1	50,2	43,337	47,101	583,947
Total	309,2	442,6	442,592	108,392	3034,85

Scanned with OKEN Scanner

În cazul unei legături simple liniare raportul de corelaţie se determină pe baza formulei (14.47). Pentru explicitarea expresiilor celor două dispersii $\sigma^2_{yy_x}$ şi σ^2_y se poate porni de la relaţia care există între valorile empirice şi valorile teoretice ale variabilei Y:

$$y_i - y_{x_i} = y_i - a - bx_i \tag{14.48}$$

Înmulţind ambii membri ai relaţiei (14.48) cu $y_i - y_{x_i}$ şi însumând rezultă:

$$\sum_{i=1}^{n}(y_i - y_{x_i})^2 = \sum_{i=1}^{n}(y_i - a - bx_i)(y_i - y_{x_i}) =$$

$$= \sum_{i=1}^{n}(y_i - y_{x_i})y_i - a\sum_{i=1}^{n}(y_i - y_{x_i}) - b\sum_{i=1}^{n}(y_i - y_{x_i})x_i$$

Dar $\sum_{i=1}^{n}(y_i - y_{x_i}) = 0$, deoarece reprezintă tocmai derivata întâi a sumei de pătrate S în raport cu a , $\sum_{i=1}^{n}(y_i - y_{x_i})x_i = 0$ fiind derivata întâi a sumei de pătrate S în raport cu b .

Deci

$$\sum_{i=1}^{n}(y_i - y_{x_i})^2 = \sum_{i=1}^{n}(y_i - y_{x_i})y_i \tag{14.49}$$

$$\sum_{i=1}^{n}(y_i - y_{x_i})y_i = \sum_{i=1}^{n}(y_i - a - bx_i)y_i$$

$$\sum_{i=1}^{n}(y_i - y_{x_i})^2 = \sum_{i=1}^{n}y_i^2 - a\sum_{i=1}^{n}y_i - b\sum_{i=1}^{n}x_i y_i \tag{14.50}$$

Iar expresia de la numitorul raportului de corelaţie (14.47) devine prin transformări elementare:

Scanned with OKEN Scanner

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-2y_i\bar{y}+\overline{y^2})=\sum_{i=1}^{n}y_i^2-2\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}+n\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n^2}$$

deci:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n} \tag{14.51}$$

Substituind în formula raportului de corelaţie (14.46) relaţiile (14.49) şi (14.50) rezultă:

$$\eta=\sqrt{1-\frac{\sum_{i=1}^{n}(y_i-y_{x_i})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}}=\sqrt{1-\frac{\sum_{i=1}^{n}y_i^2-a\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}}}=$$

$$=\sqrt{\frac{\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}-\sum_{i=1}^{n}y_i^2+a\sum_{i=1}^{n}y_i+b\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}}}$$

deci

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} \qquad (14.52)$$

Calculul simplificat al raportului de corelaţie simplă liniară se simplifică pe baza datelor din tabelul 14.3.

Substituind în formula (14.52) elementele de calcul din tabelul 14.3, obţinem:

$$\eta = \sqrt{\frac{(-5{,}7144)\cdot 442{,}6 + 1{,}7456 \cdot 9726{,}46 - \frac{442{,}6^2}{17}}{14558{,}08 - \frac{442{,}6^2}{17}}} = \sqrt{0{,}9642} = 0{,}98$$

S-a obţinut acelaşi rezultat ca şi în cazul aplicării formulei de bază a raportului de corelaţie (14.47).

Dacă perechile de valori corelate sunt ponderate cu n_i, atunci formula raportului de corelaţie (14.52) devine:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} y_i n_i + b\sum_{i=1}^{k} x_i y_i n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} y_i^2 n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (14.53)$$

Scanned with OKEN Scanner

În cazul unor distribuţii bidimensionale (x_i, y_j, n_{ij}) se aplică o formulă modificată care să ţină seama de indicii de variaţie ai celor două variabile şi de frecvenţele perechilor de valori corelate n_{ij}:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{p} y_j n_{.j} + b\sum_{i=1}^{k}\sum_{j=1}^{p} x_i y_j n_{ij} - \frac{\left(\sum_{j=1}^{p} y_j n_{.j}\right)^2}{\sum_{j=1}^{p} n_{.j}}}{\sum_{j=1}^{p} y_j^2 n_{.j} - \frac{\left(\sum_{j=1}^{p} y_j n_{.j}\right)^2}{\sum_{j=1}^{p} n_{.j}}}} \qquad (14.54)$$

Exemplu: Aplicând formula (14.54) la exemplul din tabelul 14.7, calculele intermediare se sistematizează cu ajutorul tabelului 14.8 şi rezultă:

$$\eta = \sqrt{\frac{30{,}066 \cdot 4525 + 3{,}7043 \cdot 127875 - \frac{4525^2}{39}}{466875 - \frac{4525^2}{39}}} = 0{,}84$$

Rezultă:

$$\eta = |r| = 0{,}84$$

Coeficientul de corelaţie simplă liniară r este egal cu rădăcina pătrată din *coeficientul de determinaţie* η_d^2:

$$\eta_d^2 = r^2 \qquad \text{şi} \qquad |r| = \sqrt{\eta_d^2} \qquad (14.55)$$

Pentru demonstrarea egalităţii (14.55), pornim de la relaţia (14.46) în care, după transformări elementare, numărătorul devine:

$$\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2=\sum_{i=1}^{n}y_{x_i}^2-2\bar{y}\sum_{i=1}^{n}y_{x_i}+n\bar{y}^2=n\bar{y}^2+\sum_{i=1}^{n}(a+bx)^2-2\bar{y}\sum_{i=1}^{n}(a+bx)=$$

$$=\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}+na^2+b^2\sum_{i=1}^{n}x_i^2+2ab\sum_{i=1}^{n}x_i-2a\sum_{i=1}^{n}y_i-\frac{2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n}$$

Substituind în această relaţie pe $a=\dfrac{\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i}{n}$, rezultă :

$$\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2=\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}+\frac{\left(\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i\right)^2}{n}+b^2\sum_{i=1}^{n}x_i^2+$$

$$+2b\frac{\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}x_i-2\frac{\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i}{n}\sum_{i=1}^{n}y_i-\frac{2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n}$$

şi

$$n\sum_{i=1}^{n}(y_{x_i}-\bar{y})^2=\left(\sum_{i=1}^{n}y_i\right)^2+\left(\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i\right)^2+nb^2\sum_{i=1}^{n}x_i^2+$$

$$+2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i-2b^2\left(\sum_{i=1}^{n}x_i\right)^2-2\left(\sum_{i=1}^{n}y_i\right)^2+2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i-2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i=$$

$$=\left(\sum_{i=1}^{n}y_i\right)^2+\left(\sum_{i=1}^{n}y_i\right)^2+b^2\left(\sum_{i=1}^{n}x_i\right)^2-2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i+nb^2\sum_{i=1}^{n}x_i^2+$$

Scanned with OKEN Scanner

$$+2b\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i-2b^2\left(\sum_{i=1}^{n}x_i\right)^2-2\left(\sum_{i=1}^{n}y_i\right)^2=$$

$$=nb^2\sum_{i=1}^{n}x_i^2-b^2\left(\sum_{i=1}^{n}x_i\right)^2=b^2\left[n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right]$$

Efectuând calculele în expresia de la numitorul lui η_d^2 se obţine:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}y_i^2-2\bar{y}\sum_{i=1}^{n}y_i+n\bar{y}^2=\sum_{i=1}^{n}y_i^2-2\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}+\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}$$

şi

$$n\sum_{i=1}^{n}(y_i-\bar{y})^2=n\left(\sum_{i=1}^{n}y_i^2-\frac{\left(\sum_{i=1}^{n}y_i\right)^2}{n}\right)=n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2$$

deci
$$\eta_d^2=\frac{b^2\left[n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right]}{n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}$$

Considerînd $b=\dfrac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}y_i\right)^2}$ se obţine:

$$\eta_d^2 = \frac{\dfrac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\left(n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)^2}{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}$$

Deci, relaţia (14.42) este demonstrată.

14.3.3. Legătura dintre coeficienţii de regresie şi coeficientul de corelaţie

În cazul regresiei lui Y în funcţie de X, între coeficientul de regresie b_{yx} şi coeficientul de corelaţie r_{yx} există următoarea relaţie:

$$b_{yx} = r_{yx}\frac{\sigma_y}{\sigma_x} \tag{14.56}$$

de unde

$$r_{yx} = b_{yx}\frac{\sigma_x}{\sigma_y} \tag{14.57}$$

Pentru verificarea relaţiei (14.55), considerăm:

$$1\,b_{yx} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Scanned with OKEN Scanner

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n}}$$

Substituind aceste elemente în relaţia (14.57) rezultă:

$$r_{yx} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \frac{\sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{n}}}{\sqrt{\frac{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{n}}} =$$

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \frac{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}}{\sqrt{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} =$$

Scanned with OKEN Scanner

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} = r_{yx}$$

În mod analog, în cazul regresiei lui X în funcţie de Y se verifică relaţiile:

$$b_{xy} = r_{xy}\frac{\sigma_x}{\sigma_y} \tag{14.58}$$

$$r_{xy} = b_{xy}\frac{\sigma_y}{\sigma_x} \tag{14.59}$$

Din relaţiile (14.58) şi (14.59) rezultă că un coeficient de corelaţie poate fi egal cu coeficientul de regresie corespunzător, atunci când dispersiile celor două variabile corelate sunt egale:

$r_{yx} = b_{yx}$ dacă $\sigma_x = \sigma_y$ iar $r_{xy} = b_{xy} = b_{yx}$

Între coeficienţii de regresie b_{yx}, b_{xy} şi coeficientul de corelaţie r există următoarea relaţie:

$$r_{xy} = \sqrt{b_{yx} \cdot b_{xy}} \tag{14.60}$$

Pentru verificarea acestei relaţii se substituie în (14.60) coeficienţii de regresie cu expresiile lor:

$$\sqrt{b_{yx} \cdot b_{xy}} = \sqrt{\frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \cdot \frac{n\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}} =$$

Scanned with OKEN Scanner

$$= \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} = r_{yx}$$

În particular, dacă cei doi coeficienţi de regresie sunt mărimi reciproce unul faţă de celălalt, cele două drepte de regresie coincid, iar coeficientul de corelaţie r va fi egal cu 1, deci corelaţia este perfectă.

Scanned with OKEN Scanner

15. REGRESIE ŞI CORELAŢIE SIMPLĂ CURBILINIE

15.1. Modele de regresie simplă curbilinie

În stabilirea expresiei analitice a legăturilor dintre fenomene, alături de modelele liniare de regresie, un loc important îl ocupă modelele unor curbe, de diferite tipuri: hiperbole, parabole, exponenţiale, logaritmice etc. Legăturile liniare, exprimate sintetic prin ecuaţia unei drepte, se caracterizează prin faptul că unei variaţii uniforme a variabilei factoriale îi corespunde o variaţie tot uniformă a variabilei rezultative. Pentru orice nivel al variabilei independente X, unei creşteri sau scăderi a acesteia cu o unitate îi corespunde o creştere sau descreştere a variabilei dependente Y cu aceeaşi mărime, egală cu coeficientul de regresie b. Legăturile de tip liniar sunt cele mai simple modele de regresie şi se întâlnesc destul de rar în cadrul fenomenelor reale din natură şi societate.

În practică, cele mai frecvente legături dintre fenomenele reale iau forma unor modele curbilinii, în care variaţia variabilei rezultative depinde şi de nivelul la care se află variabila independentă. Teoretic, se poate imagina un număr nelimitat de curbe care pot fi descrise de funcţii matematice, în realitate, însă, se utilizează un număr redus de modele de regresie curbilinie şi anume:

1. *Modele parabolice:*

- parabola de gradul doi $y_x = a + bx + cx^2$ (15.1)

- parabola de gradul trei $y_x = a + bx + cx^2 + dx^3$ (15.2)

- parabola semilogaritmică de gradul doi $lg\, y_x = a + bx + cx^2$ (15.3)

sau

$$y_x = a + b\, lg\, x + c(lg\, x)^2 \quad (15.4)$$

- parabola semilogaritmică de gradul trei

$$lg\, y_x = a + bx + cx^2 + dx^3 \quad (15.5)$$

Scanned with OKEN Scanner

sau $$y_x = a + b\lg x + c(\lg x)^2 + d(\lg x)^3 \quad (15.6)$$

• parabola dublu logaritmică de gradul doi

$$\lg y_x = a + b\lg x + c(\lg x)^2 \quad (15.7)$$

• parabola dublu logaritmică de gradul trei

$$\lg y_x = a + b\lg x + c(\lg x)^2 + d(\lg x)^3 \quad (15.8)$$

2. *Modele hiperbolice*

$$y_x = a + \frac{b}{x} \quad (15.9)$$

sau

$$y_x = \frac{1}{a+bx}, \quad \frac{1}{y_x} = a + bx, \quad y_x = \frac{x}{ax+b} \quad (15.10)$$

3. *Modele de tip exponenţial:* $y_x = ae^{bx}$ (15.11)

$$y_x = ae^{\frac{b}{x}} \quad (15.12)$$

$$y_x = \frac{1}{a + be^{-x}} \quad (15.13)$$

4. *Modelul unei funcţii putere* $y_x = ax^b$ (15.14)

5. *Modelul unei funcţii de tip logaritmic* $y_x = a + b\lg x$ (15.15)

15.1.1. Legături de tip parabolic

Parabola de gradul doi $y_x = a + bx + cx^2$ a cărei expresie grafică se prezintă în figura 15.1, are o serie de particularităţi a căror cunoaştere poate să orienteze mai bine cercetătorul în alegerea acestui tip de funcţie drept model de regresie.

Aşa cum se observă din figura 15.1 şi figura 15.2 parabola de gradul doi obişnuită şi cea semilogaritmică prezintă un punct de extrem, de maxim sau de minim, în funcţie de semnul coeficientului de regresie c. Dacă $c < 0$ atunci curba

Scanned with OKEN Scanner

abscisei punctului de extrem este deosebit de importantă în activitatea practică, deoarece reprezintă nivelul variabilei factoriale care corespunde nivelului maxim sau celui minim al variabilei rezultative.

Abscisa punctului de extrem al parabolei de gradul doi se obţine, după cum se ştie, egalând cu zero derivata de ordinul întâi, astfel:

$$y' = b + 2cx = 0$$

de unde

$$x = -\frac{b}{2c} \qquad (15.17)$$

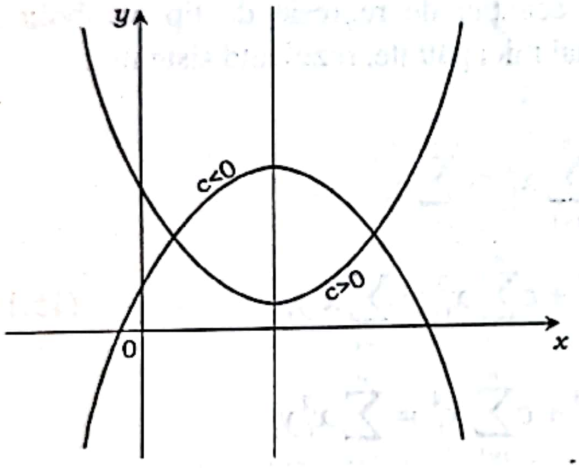


Figura15.1. Parabola de gradul doi $y = a + bx + cx^2$

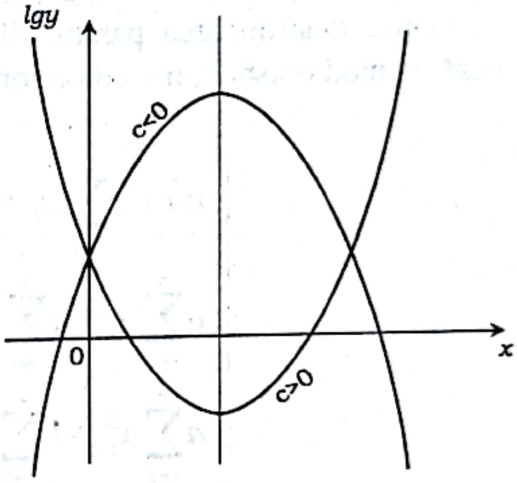


Figura15.2. Parabola semilogaritmică de gradul doi $\lg y = a + bx + cx^2$

Desigur, în practică, legăturile de tip parabolic se vor prezenta sub forma unor segmente ale curbei crescătoare sau descrescătoare, situate de regulă în cadranul întâi al axelor de coordonate, deoarece indicatorii sociali şi economici reali au valori pozitive.

Indiferent de forma ei, concavă sau convexă, parabola de gradul doi este o curbă simetrică faţă de ordonata punctului de extrem.

Ca şi în cazul dreptei de regresie, parametrii parabolei de gradul doi se pot determina cu ajutorul metodei punctelor selecţionate sau cu ajutorul metodei celor mai mici pătrate.

Scanned with OKEN Scanner

Întrucât parabola de gradul doi depinde de trei parametri se vor selecta trei puncte echidistante şi se va obţine un sistem format din trei ecuaţii. Punctele selectate se vor situa câte unul la fiecare capăt al segmentului parabolei şi unul la mijlocul ei.

Dacă punctele selectate sunt de coordonate (x_1, y_1), (x_2, y_2), (x_3, y_3) atunci se poate scrie sistemul:

$$\begin{cases} a + bx_1 + cx_1^2 = y_1 \\ a + bx_2 + cx_2^2 = y_2 \\ a + bx_3 + cx_3^2 = y_3 \end{cases}$$

Rezolvând acest sistem se vor obţine valorile celor trei necunoscute, care sunt tocmai parametrii a, b, şi c ai parabolei de gradul doi.

Pentru determinarea parametrilor ecuaţiei de regresie de tip parabolic se utilizează, în mod obişnuit, metoda celor mai mici pătrate, rezultând sistemul:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 y_i \end{cases} \qquad (15.18)$$

Dacă frecvenţa perechilor de valori corelate (x_i, y_i) este diferită de unitate şi egală cu n_i, atunci sistemul (15.18) devine:

$$\begin{cases} a\sum_{i=1}^{k} n_i + b\sum_{i=1}^{k} x_i n_i + c\sum_{i=1}^{k} x_i^2 n_i = \sum_{i=1}^{k} y_i n_i \\ a\sum_{i=1}^{k} x_i n_i + b\sum_{i=1}^{k} x_i^2 n_i + c\sum_{i=1}^{k} x_i^3 n_i = \sum_{i=1}^{k} x_i y_i n_i \\ a\sum_{i=1}^{k} x_i^2 n_i + b\sum_{i=1}^{k} x_i^3 n_i + c\sum_{i=1}^{k} x_i^4 n_i = \sum_{i=1}^{k} x_i^2 y_i n_i \end{cases} \qquad (15.19)$$

Scanned with OKEN Scanner

Dacă datele iniţiale se prezintă sub forma unui tabel de corelaţie (x_i, y_i, n_{ij}), sistemul (15.18) devine:

$$\begin{cases} a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} x_i n_{i.} + c\sum_{i=1}^{s} x_i^2 n_{i.} = \sum_{j=1}^{k} y_j n_{.j} \\ a\sum_{i=1}^{s} x_i n_{i.} + b\sum_{i=1}^{s} x_i^2 n_{i.} + c\sum_{i=1}^{s} x_i^3 n_{i.} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} \\ a\sum_{i=1}^{s} x_i^2 n_{i.} + b\sum_{i=1}^{s} x_i^3 n_{i.} + c\sum_{i=1}^{s} x_i^4 n_{i.} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i^2 y_j n_{ij} \end{cases} \quad (15.20)$$

Intensitatea corelaţiei simple parabolice se măsoară cu ajutorul raportului de corelaţie, definit de relaţia:

$$\eta = \sqrt{\frac{\sigma_y^2 - \sigma_{yy_x}^2}{\sigma_y^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - y_x)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

sau

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i^2 y_i - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} \quad (15.21)$$

Dacă perechile de valori empirice (x_i, y_i) ale celor două variabile corelate sunt ponderate cu frecvenţele n_i, atunci în formula raportului de corelaţie (15.21) va trebui ca fiecare variabilă să fie ponderată cu frecvenţa corespunzătoare, astfel:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} y_i n_i + b\sum_{i=1}^{k} x_i y_i n_i + c\sum_{i=1}^{k} x_i^2 y_i n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} y_i^2 n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.22)$$

Atunci când datele iniţiale, cu privire la cele două variabile corelate, se prezintă
sub forma unui tabel de corelaţie, formula raportului de corelaţie simplă parabolică devine:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} y_j n_{.j} + b\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} + c\sum_{i=1}^{s}\sum_{j=1}^{k} x_i^2 y_j n_{ij} - \frac{\left(\sum_{j=1}^{k} y_j n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{j=1}^{k} y_j^2 n_{.j} - \frac{\left(\sum_{j=1}^{k} y_j n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.23)$$

Să considerăm, pentru exemplificarea regresiei şi corelaţiei simple parabolice, legătura dintre valoarea fondurilor fixe şi valoarea producţiei globale industriale pe judeţele *României* în anul 1976. Datele cu privire la aceste două variabile se prezintă în tabelul 15.1.

S-a exclus din calcul *Municipiul Bucureşti*, deoarece acesta prezintă caracteristici specifice şi nu poate fi asimilat cu judeţele ţării.

Pentru a obţine o primă imagine asupra formei legăturii dintre valoarea fondurilor fixe şi valoarea producţiei globale pe judeţe, se reprezintă grafic această distribuţie bidimensională (fig. 15.4).

Scanned with OKEN Scanner

Tabelul 15.1. Distribuţia judeţelor după valoarea fondurilor fixe şi valoarea producţiei globale în anul 1976 (miliarde lei)

Y / X	- 10,0	10,1 - 20	20,1 - 30	30,1 - 40	40,1 -	Total
- 10	6	-	-	-	-	6
10,1 - 20	7	3	-	-	-	10
20,1 - 30	4	7	2	-	-	13
30,1 - 40	-	2	2	-	-	4
40,1 -	-	1	3	1	1	6
Total	17	13	7	1	1	39

Sursa: Prelucrat după *Anuarul statistic al României*, 1977, p.93, 117

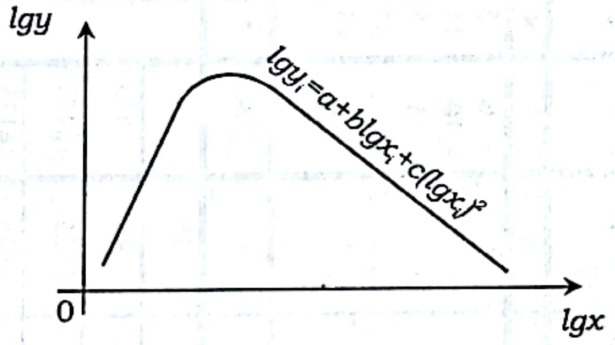


Figura 15.4. Legătura dintre valoarea fondurilor fixe şi valoarea producţiei globale pe judeţe în anul 1976

Se observă din figura 15.4 că legătura dintre variabile se apropie de o parabolă de gradul doi.

Pentru determinarea parametrilor parabolei de gradul doi, pe baza datelor din tabelul 15.1, calculele preliminare se sistematizează cu ajutorul tabelului 15.2.

Simplificarea calculelor se face considerând valorile reduse x_i' şi y_i' ale celor două variabile, obţinute prin schimbarea originii de calcul, astfel:

$$x_i' = \frac{x_i - x_0}{k_x} \quad \text{şi} \quad y_i' = \frac{y_j - y_0}{k_y}$$

Substituind elementele de calcul din tabelul 15.2 în sistemul (15.20), rezultă:

Scanned with OKEN Scanner

Tabelul 15.2. Calculele preliminare pentru determinarea indicatorilor de regresie si corelaţie simplă parabolică

Y / X	-10,0	10,1-20,0	20,1-30,0	30,1-40,0	40,1-	Total $n_{i.}$	x_i	$x_i n_{i.}$	$x_i^2 n_{i.}$	$x_i^3 n_{i.}$	$\sum_{j=1}^{k} x_i y_j n_{ij}$	$\sum_{j=1}^{k} x_i^2 y_j n_{ij}$	$x_i^4 n_{i.}$
1	2	3	4	5	6	7	8	9	10	11	12	13	14
-10,0	6	-	-	-	-	6	5	30	150	750	150	750	3750
10,1 - 20,0	7	3	-	-	-	10	15	150	2250	33750	1200	18000	506250
20,1 - 30,0	4	7	2	-	-	13	25	325	8125	203125	4375	109375	5078125
30,1 - 40,0	-	2	2	-	-	4	35	140	4900	171500	2800	98000	6002500
40,1 -	-	1	3	1	1	6	45	270	12150	1761750	7650	344250	79278750
Total $n_{.j}$	17	13	7	1	1	39	-	915	27575	2170875	16275	570375	90869375
y_j	5	15	25	35	45	-							
$y_j n_{.j}$	85	195	175	35	45	535							
$y_j^2 n_{.j}$	425	2925	4375	1225	2025	10975							
$\sum_{i=1}^{s} x_i y_j n_{ij}$	1175	5025	6375	1575	2025	16175							
$\sum_{i=1}^{s} x_i^2 y_j n_{ij}$	21125	142875	244375	70875	91125	570375							

Scanned with OKEN Scanner

Tabelul 15.2. Calculele preliminare pentru determinarea indicatorilor de regresie si corelaţie simplă parabolică

Y / X	-10,0	10,1-20,0	20,1-30,0	30,1-40,0	40,1-	Total $n_{i.}$	x_i	$x_i n_{i.}$	$x_i^2 n_{i.}$	$x_i^3 n_{i.}$	$\sum_{j=1}^{k} x_i y_j n_{ij}$	$\sum_{j=1}^{k} x_i^2 y_j n_{ij}$	$x_i^4 n_{i.}$
1	2	3	4	5	6	7	8	9	10	11	12	13	14
-10,0	6	-	-	-	-	6	5	30	150	750	150	750	3750
10,1 - 20,0	7	3	-	-	-	10	15	150	2250	33750	1200	18000	506250
20,1 - 30,0	4	7	2	-	-	13	25	325	8125	203125	4375	109375	5078125
30,1 - 40,0	-	2	2	-	-	4	35	140	4900	171500	2800	98000	6002500
40,1 -	-	1	3	1	1	6	45	270	12150	1761750	7650	344250	79278750
Total $n_{.j}$	17	13	7	1	1	39	-	915	27575	2170875	16275	570375	90869375
y_j	5	15	25	35	45	-							
$y_j n_{.j}$	85	195	175	35	45	535							
$y_j^2 n_{.j}$	425	2925	4375	1225	2025	10975							
$\sum_{i=1}^{s} x_i y_j n_{ij}$	1175	5025	6375	1575	2025	16175							
$\sum_{i=1}^{s} x_i^2 y_j n_{ij}$	21125	142875	244375	70875	91125	570375							

Scanned with OKEN Scanner

Tabelul 15.3. Rezolvarea sistemului de ecuaţii normale prin metoda lui *Gauss*

	a	b	c	
a	39	915	27575	535
b	915	27575	2170875	16175
c	27575	2170875	90869375	570375
a	1	23,4616	707,0513	13,7180
b	0	6107,6653	1523922,2624	3623,0612
c	0	1523922,2624	71372435,8989	192102,5642
a	1	0	5146,8478	-0,1995
b	0	1	249,5098	0,5932
c	0	0	308961103,0073	-711886,7110
a	1	0	0	-12,0372
b	0	1	0	1,1670
c	0	0	1	-0,0023

$$39a + 915b + 27575c = 535$$
$$915a + 27575b + 2170875c = 16175$$
$$27575a + 2170875b + 90869375c = 570375$$

Sistemul obţinut se poate rezolva prin metoda lui *Gauss* (tabelul 15.3).

Substituind parametrii *a*, *b* şi *c* de pe ultima coloană şi din ultima iteraţie a tabelului 15.3, în ecuaţia parabolei de grad doi (15.1), se obţine ecuaţia de regresie a producţiei industriale, în funcţie de valoarea fondurilor fixe pe judeţe:

$$y_{x_i} = -12{,}0372 + 1{,}1670x_i - 0{,}0023x_i^2$$

Înlocuind succesiv valorile lui x_i în această ecuaţie, rezultă:

$x_1 = 5$ $\quad y_1 = -6{,}2597$

$x_2 = 15$ $\quad y_2 = 4{,}9503$

$x_3 = 25$ $\quad y_3 = 5{,}7003$

$x_4 = 35$ $\quad y_4 = 24{,}1503$

$x_5 = 45$ $\quad y_5 = 35{,}8203$

Scanned with OKEN Scanner

Reprezentând grafic valorile teoretice y_{x_i} în figura 15.4, se obţine un arc de parabolă care trece prin centrul norului de puncte, ceea ce dovedeşte că ecuaţia de regresie de formă parabolică descrie în mod corespunzător legătura dintre valoarea producţiei globale şi valoarea fondurilor fixe pe judeţe.

Intensitatea legăturii dintre valoarea producţiei globale industriale şi valoarea fondurilor fixe pe judeţe se poate determina cu ajutorul relaţiei (15.27), astfel:

$$\eta = \sqrt{\frac{-12{,}0372 \cdot 535 + 1{,}167 \cdot 16175 - 0{,}0023 \cdot 570375 - \frac{535^2}{39}}{10975 - \frac{535^2}{39}}} = \sqrt{0{,}8346} = 0{,}91$$

iar $$\eta^2 = 83{,}46\%$$

Aceasta înseamnă că între gradul de înzestrare tehnică a judeţelor şi valoarea producţiei globale există o legătură parabolică intensă şi anume, în medie 83,46% din valoarea producţiei globale industriale este determinată de valoarea fondurilor fixe pe judeţe şi numai 16,54% se datorează altor factori.

În cazul parabolei de gradul doi semilogaritmică (15.3), reprezentată grafic în figura 15.2, prin aplicarea metodei celor mai mici pătrate rezultă următoarele forme ale sistemului de ecuaţii normale:

- când perechile de valori corelate $x_i, \lg y_i$ sunt neponderate:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} lg\, y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i\, lg\, y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2\, lg\, y_i \end{cases} \qquad (15.24)$$

- când perechile de valori corelate $x_i, \lg y_i$ sunt ponderate cu frecvenţa n_i :

Scanned with OKEN Scanner

$$\begin{cases} a\sum_{i=1}^{k} n_i + b\sum_{i=1}^{k} x_i n_i + c\sum_{i=1}^{k} x_i^2 n_i = \sum_{i=1}^{k} (\lg y_i) n_i \\ a\sum_{i=1}^{k} x_i n_i + b\sum_{i=1}^{k} x_i^2 n_i + c\sum_{i=1}^{k} x_i^3 n_i = \sum_{i=1}^{k} x_i (\lg y_i) n_i \\ a\sum_{i=1}^{k} x_i^2 n_i + b\sum_{i=1}^{k} x_i^3 n_i + c\sum_{i=1}^{k} x_i^4 n_i = \sum_{i=1}^{k} x_i^2 (\lg y_i) n_i \end{cases} \quad (15.25)$$

• când valorile celor două variabile corelate X şi Y se prezintă sub forma unui tabel de corelaţie:

$$\begin{cases} a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} x_i n_{i.} + c\sum_{i=1}^{s} x_i^2 n_{i.} = \sum_{j=1}^{k} (\lg y_j) n_{.j} \\ a\sum_{i=1}^{s} x_i n_{i.} + b\sum_{i=1}^{s} x_i^2 n_{i.} + c\sum_{i=1}^{s} x_i^3 n_{i.} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i (\lg y_j) n_{ij} \\ a\sum_{i=1}^{s} x_i^2 n_{i.} + b\sum_{i=1}^{s} x_i^3 n_{i.} + c\sum_{i=1}^{s} x_i^4 n_{i.} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i^2 (\lg y_j) n_{ij} \end{cases} \quad (15.26)$$

Formulele de calcul simplificat al raportului de corelaţie în acest caz vor fi:

• în cazul perechilor de valori neponderate $(x_i, \lg y_i)$:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} \lg y_i + b\sum_{i=1}^{n} x_i \lg y_i + c\sum_{i=1}^{n} x_i^2 \lg y_i - \frac{\left(\sum_{i=1}^{n} \lg y_i\right)^2}{n}}{\sum_{i=1}^{n} (\lg y_i)^2 - \frac{\left(\sum_{i=1}^{n} \lg y_i\right)^2}{n}}} \quad (15.27)$$

• în cazul perechilor de valori $(x_i, \lg y_i)$ ponderate cu n_i:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} n_i \lg y_i + b\sum_{i=1}^{k} n_i x_i \lg y_i + c\sum_{i=1}^{k} n_i x_i^2 \lg y_i - \frac{\left(\sum_{i=1}^{k} n_i \lg y_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} n_i (\lg y_i)^2 - \frac{\left(\sum_{i=1}^{k} n_i \lg y_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.28)$$

- în cazul prezentării celor două variabile sub forma unui tabel de corelaţie:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} n_{.j} \lg y_j + b\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i \lg y_j + c\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i^2 \lg y_j - \frac{\left(\sum_{j=1}^{k} n_{.j} \lg y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{j=1}^{k} n_{.j} (\lg y_j)^2 - \frac{\left(\sum_{j=1}^{k} n_{.j} \lg y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.29)$$

Legătura simplă dintre fenomene poate îmbrăca şi forma unei parabole de gradul trei (15.2).

Parabola de gradul trei este o curbă asimetrică care prezintă două puncte de extrem unul de maxim şi altul de minim (fig. 15.3) şi un singur punct de inflexiune, ea corespunde situaţiei în care derivata întâi are două rădăcini reale şi diferite:

$$y' = b + 2cx + 3dx^2$$

iar $$\Delta = c^2 - 3db > 0$$

Scanned with OKEN Scanner

Parabola de gradul trei depinde de patru parametri a, b, c, d şi va fi, deci, necesar să se selecteze patru puncte echidistante pentru estimarea celor patru parametri ai curbei.

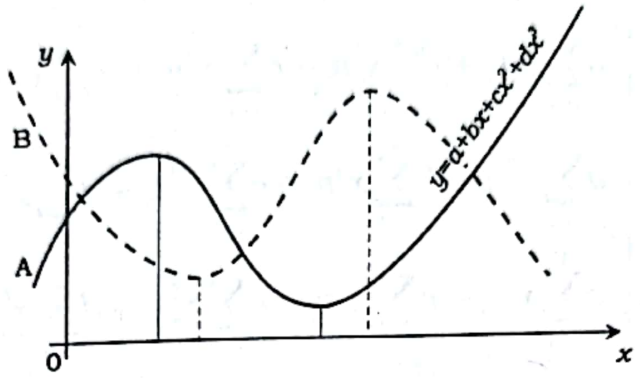


Figura 15.3. Parabola de gradul trei $y_x = a + bx + cx^2 + dx^3$

Când $d > 0$ atunci apare întâi punctul de maxim şi apoi cel de minim (fig. 15.3 - curba A), iar când $d < 0$ punctul de minim precede pe cel de maxim (fig. 15.3 - curba B).

Parametrii parabolei de gradul trei se pot obţine şi cu ajutorul metodei celor mai mici pătrate prin rezolvarea sistemului de ecuaţii normale:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i + c\sum_{i=1}^{n} x_i^2 + d\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i^3 + d\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^4 + d\sum_{i=1}^{n} x_i^5 = \sum_{i=1}^{n} x_i^2 y_i \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^4 + c\sum_{i=1}^{n} x_i^5 + d\sum_{i=1}^{n} x_i^6 = \sum_{i=1}^{n} x_i^3 y_i \end{cases} \qquad (15.30)$$

Scanned with OKEN Scanner

În cazul în care frecvenţele n_i ale perechilor de valori x_i, y_i sunt diferite de unitate, sistemul (15.30) devine:

$$\begin{cases} a\sum_{i=1}^{k} n_i + b\sum_{i=1}^{k} x_i n_i + c\sum_{i=1}^{k} x_i^2 n_i + d\sum_{i=1}^{k} x_i^3 n_i = \sum_{i=1}^{k} y_i n_i \\ a\sum_{i=1}^{k} x_i n_i + b\sum_{i=1}^{k} x_i^2 n_i + c\sum_{i=1}^{k} x_i^3 n_i + d\sum_{i=1}^{k} x_i^4 n_i = \sum_{i=1}^{k} x_i y_i n_i \\ a\sum_{i=1}^{k} x_i^2 n_i + b\sum_{i=1}^{k} x_i^3 n_i + c\sum_{i=1}^{k} x_i^4 n_i + d\sum_{i=1}^{k} x_i^5 n_i = \sum_{i=1}^{k} x_i^2 y_i n_i \\ a\sum_{i=1}^{k} x_i^3 n_i + b\sum_{i=1}^{k} x_i^4 n_i + c\sum_{i=1}^{k} x_i^5 n_i + d\sum_{i=1}^{k} x_i^6 n_i = \sum_{i=1}^{k} x_i^3 y_i n_i \end{cases}$$

Dacă variantele variabilelor corelate se prezintă sub forma unui tabel de corelaţie (x_i, y_j, n_{ij}), atunci sistemul de ecuaţii normale ia forma:

$$\begin{cases} a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} x_i n_{i\cdot} + c\sum_{i=1}^{s} x_i^2 n_{i\cdot} + d\sum_{i=1}^{s} x_i^3 n_{i\cdot} = \sum_{j=1}^{k} y_j n_{\cdot j} \\ a\sum_{i=1}^{s} x_i n_{i\cdot} + b\sum_{i=1}^{s} x_i^2 n_{i\cdot} + c\sum_{i=1}^{s} x_i^3 n_{i\cdot} + d\sum_{i=1}^{s} x_i^4 n_{i\cdot} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} \\ a\sum_{i=1}^{s} x_i^2 n_{i\cdot} + b\sum_{i=1}^{s} x_i^3 n_{i\cdot} + c\sum_{i=1}^{s} x_i^4 n_{i\cdot} + d\sum_{i=1}^{s} x_i^5 n_{i\cdot} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i^2 y_j n_{ij} \\ a\sum_{i=1}^{s} x_i^3 n_{i\cdot} + b\sum_{i=1}^{s} x_i^4 n_{i\cdot} + c\sum_{i=1}^{s} x_i^5 n_{i\cdot} + d\sum_{i=1}^{s} x_i^6 n_{i\cdot} = \sum_{i=1}^{s}\sum_{j=1}^{k} x_i^3 y_j n_{ij} \end{cases} \quad (15.31)$$

Prin rezolvarea acestor sisteme, cu ajutorul uneia dintre metodele cunoscute, se obţin valorile celor patru parametri ai ecuaţiei de regresie în fiecare dintre cele trei modalităţi posibile de prezentare a datelor statistice.

Intensitatea legăturii dintre variabile (x_i, y_i) se măsoară cu ajutorul raportului de corelaţie dat de formula:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} x_i y_i + c\sum_{i=1}^{n} x_i^2 y_i + d\sum_{i=1}^{n} x_i^3 y_i - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} \qquad (15.32)$$

În cazul ponderării perechilor de valori ale variabilelor cu frecvenţa (x_i, y_i, n_i), raportul de corelaţie are următoarea expresie:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} y_i n_i + b\sum_{i=1}^{k} x_i y_i n_i + c\sum_{i=1}^{k} x_i^2 y_i n_i + d\sum_{i=1}^{k} x_i^3 y_i n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} y_i^2 n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.33)$$

Dacă cele două variabile corelate X şi Y se prezintă sub forma unui tabel de corelaţie, raportul de corelaţie devine:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} y_j n_{.j} + b\sum_{i=1}^{s}\sum_{j=1}^{k} x_i y_j n_{ij} + c\sum_{i=1}^{s}\sum_{j=1}^{k} x_i^2 y_j n_{ij} + d\sum_{i=1}^{s}\sum_{j=1}^{k} x_i^3 y_j n_{ij} - \frac{\left(\sum_{j=1}^{k} y_j n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{j=1}^{k} y_j^2 n_{.j} - \frac{\left(\sum_{j=1}^{k} y_i n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.34)$$

Desigur, şi pentru parabolele de grad superior indicatorii de regresie şi corelaţie simplă se vor calcula şi analiza în mod asemănător. Cu cât creşte gradul parabolei, cu atât creşte şi numărul punctelor de extrem. Astfel, în cazul unei drepte (care poate fi considerată drept o parabolă de gradul întâi), aceasta nu are nici un punct de extrem şi depinde de doi parametri. Parabola de gradul doi depinde de trei parametri şi are un punct de extrem. În general, parabola de gradul *n* depinde de *n+1* parametri şi are *n-1* puncte de extrem şi *n-2* puncte de inflexiune. În practica statistică, se tinde a se folosi curbe de regresie cât mai simple şi de aceea numai în cazuri rare se recurge la parabole de un grad mai mare decât trei.

15.1.2. Legături de tip hiperbolic

Unele legături simple dintre fenomenele reale pot să ia forma unei hiperbole. Ecuaţia hiperbolei poate lua mai multe expresii în funcţie de caracteristicile ei principale, astfel:

$$y = a + \frac{b}{x} \quad \text{sau} \quad y = \frac{ax + b}{x} \qquad (15.35)$$

unde $(0 < x < \infty)$

Scanned with OKEN Scanner

Graficul funcţiei hiperbolice (15.35) se prezintă în figura 15.5.

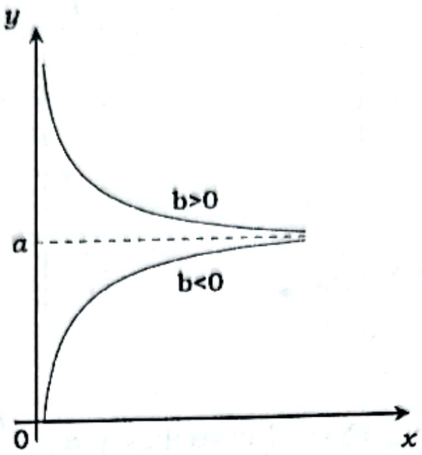


Figura 15.5. Hiperbola $y = a + \frac{b}{x}$

Curba descrisă de ecuaţia (15.35) are două asimptote: (vezi figura 15.5) una orizontală $y = a$ şi una verticală $x = 0$. Pentru $b > 0$, curba este descrescătoare şi tinde asimptotic către a, iar dacă $b < 0$, atunci hiperbola este descrisă de o curbă crescătoare care tinde asimptotic spre o valoare limită superioară a, când x tinde către $+\infty$. Când $b < 0$, legătura dintre variabile este directă sau pozitivă, iar când $b > 0$, legătura dintre variabile este inversă sau negativă.

Dependenţa de tip hiperbolic poate lua şi o altă formă şi anume:

$$y_x = \frac{1}{a + bx} \quad unde \ (-\frac{a}{b} < x < \infty) \qquad (15.36)$$

Graficul acestei funcţii se reprezintă în figura 15.6.

După cum se observă din fig. 15.6, cele două curbe hiperbolice au tot două asimptote: una orizontală $y = 0$ şi alta verticală $x = -\frac{a}{b}$. În aplicaţiile practice se întâlneşte doar cazul A şi numai partea din curbă situată în cadranul întâi, corespunzătoare unor valori pozitive ale celor două variabile. Curba din figura 15.6. este o curbă descrescătoare ce corespunde unor legături inverse între variabilele corelate.

Scanned with OKEN Scanner

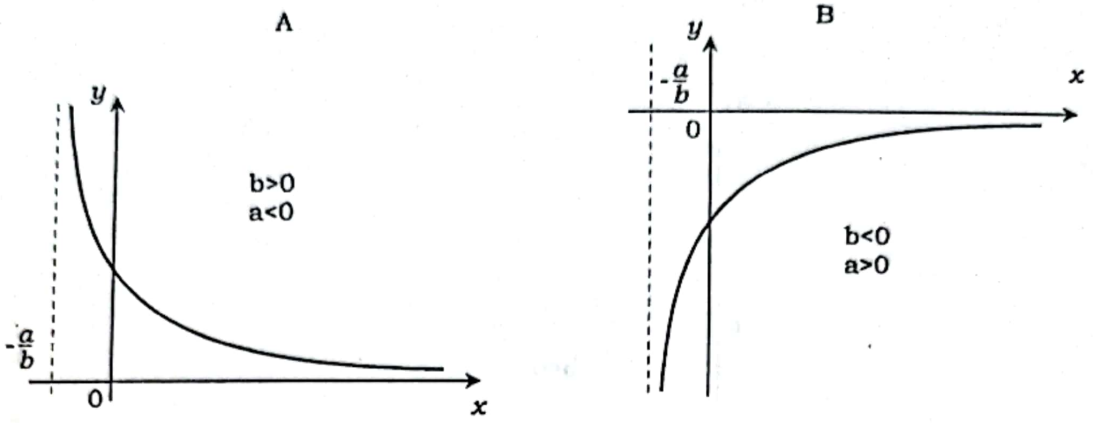


Figura 15.6. Hiperbola $y = \frac{1}{a+bx}$

Dependenţa de tip hiperbolic poate fi descrisă şi de ecuaţia:

$$y = \frac{1}{a+\frac{b}{x}} \quad unde \quad x \in (-\frac{b}{a}, \infty) \setminus \{0\} \tag{15.37}$$

sau $$y = \frac{x}{ax+b} \quad unde \quad (-\frac{b}{a} < x < \infty)$$

Hiperbola din figura 15.7 are două asimptote: una orizontală $y = \frac{1}{a}$ şi una verticală $x = -\frac{b}{a}$. Curba din fig. 15.7. *A* prezintă o legătură inversă, iar cea din fig. 15.7. *B* o legătură directă între variabilele corelate.

Parametrii ecuaţiei de regresie de tip hiperbolic (15.35) se determină cu ajutorul metodei celor mai mici pătrate din sistemul:

$$\begin{cases} na + b\sum_{i=1}^{n}\frac{1}{x_i} = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n}\frac{1}{x_i} + b\sum_{i=1}^{n}\frac{1}{x_i^2} = \sum_{i=1}^{n}\frac{y_i}{x_i} \end{cases} \tag{15.38}$$

Scanned with OKEN Scanner

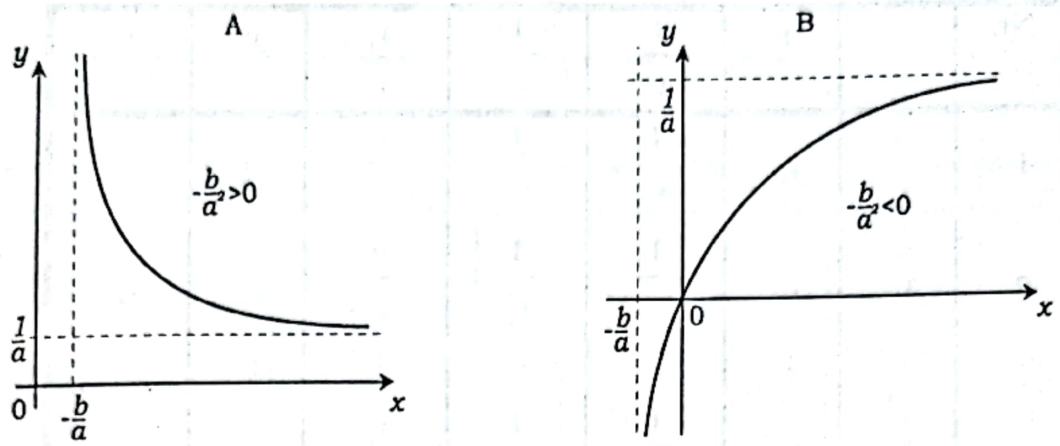


Figura 15.7. Hiperbola $y = \dfrac{1}{a + \dfrac{b}{x}}$

Aplicând regula lui *Cramer*, pentru rezolvarea sistemului rezultă formulele uzuale de calcul ale parametrilor ecuaţiei de regresie de tip hiperbolic:

$$a = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} \frac{1}{x_i^2} - \sum_{i=1}^{n} \frac{1}{x_i} \sum_{i=1}^{n} \frac{y_i}{x_i}}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^2} \tag{15.39}$$

$$b = \frac{n \sum_{i=1}^{n} \frac{y_i}{x_i} - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} \frac{1}{x_i}}{n \sum_{i=1}^{n} \frac{1}{x_i^2} - \left(\sum_{i=1}^{n} \frac{1}{x_i} \right)^2} \tag{15.40}$$

Calculele preliminare ale coeficienţilor parametrilor a şi b se pot sistematiza cu ajutorul tabelului 15.4.

Formulele parametrilor ecuaţiei de regresie de tip hiperbolic (15.39) şi (15.40) se utilizează în practica statistică atunci când valorile variabilelor corelate X şi Y se prezintă sub forma a două şiruri paralele.

Scanned with OKEN Scanner

Tabelul 15.4

Nr. crt.	X	Y	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\frac{y_i}{x_i}$	y_i^2	y_{x_i}
1	x_1	y_1	$\frac{1}{x_1}$	$\frac{1}{x_1^2}$	$\frac{y_1}{x_1}$	y_1^2	y_{x_1}
2	x_2	y_2	$\frac{1}{x_2}$	$\frac{1}{x_2^2}$	$\frac{y_2}{x_2}$	y_2^2	y_{x_2}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	x_i	y_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	$\frac{y_i}{x_i}$	y_i^2	y_{x_i}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	x_n	y_n	$\frac{1}{x_n}$	$\frac{1}{x_n^2}$	$\frac{y_n}{x_n}$	y_n^2	y_{x_n}
Total	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} \frac{1}{x_i}$	$\sum_{i=1}^{n} \frac{1}{x_i^2}$	$\sum_{i=1}^{n} \frac{y_i}{x_i}$	$\sum_{i=1}^{n} y_i^2$	$\sum_{i=1}^{n} y_{x_i}$

Dacă perechile de valori corelate sunt ponderate cu frecvenţa (x_i, y_i, n_i), atunci sistemul de ecuaţii normale (15.38) devine:

$$\begin{cases} a\sum_{i=1}^{n} n_i + b\sum_{i=1}^{n} \frac{n_i}{x_i} = \sum_{i=1}^{n} y_i n_i \\ a\sum_{i=1}^{n} \frac{n_i}{x_i} + b\sum_{i=1}^{n} \frac{n_i}{x_i^2} = \sum_{i=1}^{n} \frac{y_i n_i}{x_i} \end{cases} \tag{15.41}$$

Sistematizarea calculelor preliminare se prezintă în tabelul 15.5.

Dacă datele iniţiale cu privire la variabilele X şi Y se prezintă sub forma unui tabel de corelaţie, atunci sistemul de ecuaţii normale devine:

$$a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i} = \sum_{j=1}^{k} y_j n_{\cdot j}$$

$$a\sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i} + b\sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i^2} = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{y_j}{x_i} n_{ij} \tag{15.42}$$

Scanned with OKEN Scanner

Tabelul 15.5.

Nr. crt.	X	Y	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
0	1	2	3	4	5	6	7	8	9	10
1	x_1	y_1	n_1	$\frac{1}{x_1}$	$\frac{n_1}{x_1}$	$\frac{1}{x_1^2}$	$\frac{n_1}{x_1^2}$	$y_1 n_1$	$\frac{y_1 n_1}{x_1}$	$y_1^2 n_1$
2	x_2	y_2	n_2	$\frac{1}{x_2}$	$\frac{n_2}{x_2}$	$\frac{1}{x_2^2}$	$\frac{n_2}{x_2^2}$	$y_2 n_2$	$\frac{y_2 n_2}{x_2}$	$y_2^2 n_2$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	x_i	y_i	n_i	$\frac{1}{x_i}$	$\frac{n_i}{x_i}$	$\frac{1}{x_i^2}$	$\frac{n_i}{x_i^2}$	$y_i n_i$	$\frac{y_i n_i}{x_i}$	$y_i^2 n_i$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	x_k	y_k	n_k	$\frac{1}{x_k}$	$\frac{n_k}{x_k}$	$\frac{1}{x_k^2}$	$\frac{n_k}{x_k^2}$	$y_k n_k$	$\frac{y_k n_k}{x_k}$	$y_k^2 n_k$
Total	-	-	$\sum_{i=1}^{k} n_i$	-	$\sum_{i=1}^{k} \frac{n_i}{x_i}$	-	$\sum_{i=1}^{n} \frac{n_i}{x_i^2}$	$\sum_{i=1}^{k} y_i n_i$	$\sum_{i=1}^{k} \frac{y_i n_i}{x_i}$	$\sum_{i=1}^{k} y_i^2 n_i$

Calculul coeficienţilor parametrilor a şi b din sistemul (15.42) se sistematizează cu ajutorul tabelului 15.6.

Intensitatea legăturii dintre variabile (x_i, y_i), în cazul unei legături simple curbilinii de tip hiperbolic, se poate determina cu ajutorul formulei generale a raportului de corelaţie:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} \frac{y_i}{x_i} - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} \qquad (15.43)$$

Scanned with OKEN Scanner

Dacă perechile de variante (x_i, y_i) ale celor două variabile sunt ponderate cu frecvenţa (x_i, y_i, n_i), atunci formula de calcul a raportului de corelaţie, pentru dependenţa de tip hiperbolic, devine:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} y_i n_i + b\sum_{i=1}^{k} \frac{y_i}{x_i} n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} y_i^2 n_i - \frac{\left(\sum_{i=1}^{k} y_i n_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.44)$$

Iar în cazul unui tabel de corelaţie (x_i, y_i, n_{ij}), va fi:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} y_j n_{.j} + b\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{y_j}{x_i} \cdot n_{ij} - \frac{\left(\sum_{j=1}^{k} y_j n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{j=1}^{k} y_j^2 n_{.j} - \frac{\left(\sum_{j=1}^{k} y_j n_{.j}\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.45)$$

Scanned with OKEN Scanner

Tabelul 15.6

Nr.crt.	Y \ X	x_1	x_2	$\cdots$	x_i	$\cdots$	x_s	Total $n_{.j}$	$y_j n_{.j}$	$\sum_{i=1}^{s} \frac{y_j}{x_i} n_{ij}$	$y_j^2 n_{.j}$
0	A	1	2	$\cdots$	i	$\cdots$	s	$s+1$	$s+2$	$s+3$	$s+4$
1	y_1	n_{11}	n_{21}	$\cdots$	n_{i1}	$\cdots$	n_{s1}	$n_{.1}$	$y_1 n_{.1}$	$\sum_{i=1}^{s} \frac{y_1}{x_i} n_{i1}$	$y_1^2 n_{.1}$
2	y_2	n_{12}	n_{22}	$\cdots$	n_{i2}	$\cdots$	n_{s2}	$n_{.2}$	$y_2 n_{.2}$	$\sum_{i=1}^{s} \frac{y_2}{x_i} n_{i2}$	$y_2^2 n_{.2}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
j	y_j	n_{1j}	n_{2j}	$\cdots$	n_{ij}	$\cdots$	n_{sj}	$n_{.j}$	$y_j n_{.j}$	$\sum_{i=1}^{s} \frac{y_j}{x_i} n_{ij}$	$y_j^2 n_{.j}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	y_k	n_{1k}	n_{2k}	$\cdots$	n_{ik}	$\cdots$	n_{sk}	$n_{.k}$	$y_k n_{.k}$	$\sum_{i=1}^{s} \frac{y_k}{x_i} n_{ik}$	$y_k^2 n_{.k}$
$k+1$	Total $(n_{i.})$	$n_{1.}$	$n_{2.}$	$\cdots$	$n_{i.}$	$\cdots$	$n_{s.}$	$n_{..}$	$\sum_{j=1}^{k} y_j n_{.j}$	$\sum_{i=1}^{s}\sum_{j=1}^{k} \frac{y_j}{x_i} n_{ij}$	$\sum_{j=1}^{k} y_j^2 n_{.j}$
$k+2$	$\frac{n_{i.}}{x_i}$	$\frac{n_{1.}}{x_1}$	$\frac{n_{2.}}{x_2}$	$\cdots$	$\frac{n_{i.}}{x_i}$	$\cdots$	$\frac{n_{s.}}{x_s}$	$\sum_{i=1}^{s} \frac{n_{i.}}{x_i}$			
$k+3$	$\frac{n_{i.}}{x_i^2}$	$\frac{n_{1.}}{x_1^2}$	$\frac{n_{2.}}{x_2^2}$	$\cdots$	$\frac{n_{i.}}{x_i^2}$	$\cdots$	$\frac{n_{s.}}{x_s^2}$	$\sum_{i=1}^{s} \frac{n_{i.}}{x_i^2}$			
$k+4$	$\sum_{i=1}^{s} \frac{y_j}{x_i} n_{i.}$	$\sum_{i=1}^{s} \frac{y_j}{x_1} n_{1j}$	$\sum_{i=1}^{s} \frac{y_j}{x_2} n_{2j}$	$\cdots$	$\sum_{i=1}^{s} \frac{y_j}{x_i} n_{ij}$	$\cdots$	$\sum_{i=1}^{s} \frac{y_j}{x_s} n_{sj}$	$\sum_{i=1}^{s}\sum_{i=1}^{k} \frac{y_j}{x_i} n_{ij}$			
$k+5$	y_{x_1}	y_{x_1}	y_{x_2}	$\cdots$	y_{x_i}	$\cdots$	y_{x_s}	-			

Scanned with OKEN Scanner

Pentru exemplificarea modului de calcul şi analiză statistică a legăturilor simple curbilinii, de tip hiperbolic, să considerăm dependenţa dintre volumul desfacerilor de mărfuri în 44 de unităţi comerciale şi nivelul relativ al cheltuielilor de circulaţie. Datele iniţiale se prezintă în tabelul 15.7.

Tabelul 15.7. Vânzările de mărfuri şi nivelul relativ al cheltuielilor de circulaţie

Volumul vânzărilor de mărfuri(milioane lei) / Nivelul relativ al cheltuielilor de circulaţie(%)	Până la 4	4-8	8-12	12-16	16-20	20şi peste	Total
Până la 3,6	-	-	-	-	4	5	9
3,6 - 4,2	-	-	5	3	1	2	11
4,2 - 4,8	-	5	3	-	-	-	8
4,8 - 5,4	1	4	-	-	-	-	5
5,4 - 6,0	3	2	-	-	-	-	5
6,0 şi peste	6	-	-	-	-	-	6
Total	10	11	8	3	5	7	44

Date convenţionale

Este ştiut faptul că unităţile comerciale mici au un nivel mai ridicat al cheltuielilor de circulaţie decât unităţile ce realizează un volum mai mare al vânzărilor de mărfuri. Pe măsură ce creşte volumul desfacerilor de mărfuri, se reduce în mod continuu nivelul relativ al cheltuielilor de circulaţie. La început, reducerea este mai mare, iar pe măsura creşterii volumului vânzărilor, reducerea relativă devine din ce în ce mai mică. Acest fapt apare mai clar dacă se observă modul de dispersare al frecvenţelor în cadrul tabelului 15.7. Deci, legătura dintre variabile se apropie de modelul unei hiperbole, după cum se observă din corelograma din fig. 15.8.

Un alt argument în favoarea ipotezei dependenţei hiperbolice dintre nivelul relativ al cheltuielilor de circulaţie şi volumul vânzărilor de mărfuri se obţine calculând mediile condiţionate ale variabilei Y ,în funcţie de diferite valori ale lui X .

Scanned with OKEN Scanner

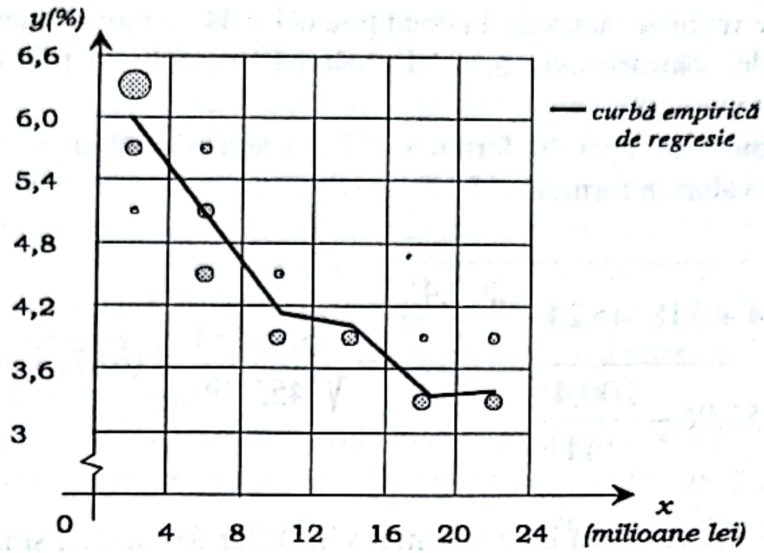


Figura 15.8. Corelaţia dintre nivelul relativ al cheltuielilor de circulaţie şi volumul vânzărilor de mărfuri

Parametrii ecuaţiei de regresie $y = a + \frac{b}{x}$ se determină cu ajutorul formulelor rezultate din rezolvarea sistemului de ecuaţii (15.42), pe baza calculelor intermediare din tabelul 15.8, astfel :

$$a = \frac{\sum_{j=1}^{k} y_j n_{\cdot j} \sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i^2} - \sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i} \sum_{i=1}^{s} \sum_{j=1}^{k} \frac{y_j}{x_i} n_{ij}}{\sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i^2} - (\sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i})^2} = \frac{200,4 \cdot 2,9308 - 8,4436 \cdot 45,23946}{44 \cdot 2,9308 - 8,4436^2} = 3,56\%$$

$$b = \frac{\sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} \sum_{j=1}^{k} \frac{y_j}{x_i} n_{ij} - \sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i} \sum_{j=1}^{k} y_j n_{\cdot j}}{\sum_{i=1}^{s} \sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i^2} - \left(\sum_{i=1}^{s} \frac{n_{i\cdot}}{x_i} \right)^2} = \frac{44 \cdot 45,23946 - 8,4436 \cdot 200,4}{44 \cdot 2,9308 - 8,4436^2} = 5,18 \; lei$$

Corelaţia dintre nivelul relativ al cheltuielilor de circulaţie şi volumul vânzărilor de mărfuri este descrisă de ecuaţia de regresie:

$$y = 3,56 + \frac{5,18}{x}$$

Scanned with OKEN Scanner

Ecuaţia de regresie arată că, în condiţiile celor 44 de unităţi comerciale luate în studiu, la 100 lei vânzări revin 3,56 lei cheltuieli variabile şi 5,18 lei cheltuieli convenţional constante.

Toate sumele ce apar în formula (15.27) sunt calculate în tabelul 15.2. Substituind aceste valori în formula (15.27), rezultă:

$$\eta = \sqrt{\frac{3{,}56 \cdot 200{,}4 + 5{,}18 \cdot 45{,}24 - \frac{200{,}4^2}{44}}{957{,}96 - \frac{200{,}4^2}{44}}} = \sqrt{\frac{35{,}0363}{45{,}229}} = \sqrt{0{,}774642} = 0{,}88$$

$\eta = 0{,}88$ înseamnă că între volumul vânzărilor de mărfuri şi nivelul relativ al cheltuielilor de circulaţie există o corelaţie strânsă.

În cazul unor legături curbilinii calculul şi analiza indicatorilor de regresie şi corelaţie pot deveni mai complicate. Pentru a evita unele calcule mai complexe şi mai greu de interpretat, în diferite situaţii concrete, se poate recurge la un *artificiu de calcul.*

Prin aplicarea acestui procedeu, funcţii complicate de regresie se reduc la o formă uşor de calculat şi apoi indicatorii de regresie şi corelaţie se determină după aceleaşi procedee ca şi în cazul legăturilor liniare. Prin liniarizare, după cum se ştie nu se schimbă nici forma reală a legăturilor curbilinii, nici esenţa indicatorilor de regresie şi corelaţie, ci doar se simplifică calculele.

Ecuaţia hiperbolei (15.35) se liniarizează astfel:

$$y' = a + b\frac{1}{x'} \tag{15.46}$$

Substituind pe $x = \frac{1}{x'}$ şi pe y' cu y rezultă:

$$y = a + bx$$

Intensitatea corelaţiei acum se poate determina, ca în cazul oricărei legături liniare, cu ajutorul coeficientului de corelaţie, astfel:

$$r_{\frac{1}{x},y} = \frac{n\sum_{i=1}^{n}\frac{1}{x_i}y_i - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n}y_i}{\sqrt{\left[n\sum_{i=1}^{n}(\frac{1}{x_i})^2 - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2\right]\left[n\sum_{i=1}^{n}y_i^2 - \left(\sum_{i=1}^{n}y_i\right)^2\right]}} \tag{15.47}$$

Scanned with OKEN Scanner

Remarcăm faptul că, atât coeficientul de corelaţie (15.47) cât şi raportul de corelaţie, calculaţi pentru aceleaşi date vor da acelaşi rezultat. S-a demonstrat în capitolul anterior că, în cazul legăturilor liniare (şi a celor liniarizate), raportul de corelaţie este identic cu valoarea absolută a coeficientului de corelaţie.

Coeficientul de corelaţie în cazul legăturilor curbilinii liniarizate se prezintă într-o formă mai convenabilă de aplicare în practică, se calculează şi se interpretează mai uşor. La fel ca şi raportul de corelaţie, oferă posibilitatea măsurării într-o formă unitară a intensităţii legăturii dintre variabile, indiferent dacă sunt liniare sau curbilinii şi se lărgesc în acest fel posibilităţile de comparabilitate a intensităţii legăturii de la un tip de ecuaţie de regresie la altul.

Desigur, formula coeficientului de corelaţie va avea expresii diferite în funcţie de modul de prezentare a variabilelor corelate. Astfel, dacă perechile de valori ale celor două variabile sunt neponderate, atunci formula de calcul este cea din (15.47). Dacă perechile de valori x_i, y_i sunt ponderate cu frecvenţele n_i, atunci coeficientul de corelaţie, pentru o legătură hiperbolică liniarizată, va fi dat de formula:

$$r_{\frac{1}{x},y} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \frac{n_i y_i}{x_i} - \sum_{i=1}^{k} \frac{n_i}{x_i} \sum_{i=1}^{k} y_i n_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \frac{n_i}{x_i^2} - \left(\sum_{i=1}^{k} \frac{n_i}{x_i}\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} y_i^2 n_i - \left(\sum_{i=1}^{k} y_i n_i\right)^2\right]}} \quad (15.48)$$

Dacă variabilele corelate hiperbolic se prezintă sub forma unui tabel de corelaţie cu dublă intrare, atunci coeficientul de corelaţie va avea formula:

$$r_{\frac{1}{x},y} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{n_{ij} y_j}{x_i} - \sum_{i=1}^{s} \frac{n_{i.}}{x_i} \sum_{j=1}^{k} n_{.j} y_j}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} \frac{n_{i.}}{x_i^2} - \left(\sum_{i=1}^{s} \frac{n_{i.}}{x_i}\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k} n_{.j} y_j^2 - \left(\sum_{j=1}^{k} n_{.j} y_j\right)^2\right]}} \quad (15.49)$$

În cazul regresiei hiperbolice, date de modelul (15.36), al cărei grafic se prezintă în figura 15.6, legătura dintre variabile poate avea expresiile:

$$y = \frac{1}{a + bx} \quad \text{sau} \quad \frac{1}{y} = a + bx$$

Scanned with OKEN Scanner

Determinarea parametrilor a şi b ai acestei funcţii cu ajutorul metodei celor mai mici pătrate se face pe baza sistemului de ecuaţii normale:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \frac{1}{y_i} \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \frac{x_i}{y_i} \end{cases} \tag{15.50}$$

Aplicând regula lui *Cramer*, pentru rezolvarea sistemului (15.50) rezultă următoarele formule de calcul al celor doi parametri:

$$a = \frac{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} \frac{1}{y_i} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \frac{x_i}{y_i}}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{15.51}$$

$$b = \frac{n\sum_{i=1}^{n} \frac{x_i}{y_i} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \frac{1}{y_i}}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} \tag{15.52}$$

Formulele (15.51) şi (15.52) sunt aplicabile pentru determinarea parametrilor ecuaţiei de regresie hiperbolică, atunci când datele iniţiale cu privire la variaţia celor două variabile se prezintă sub forma a două şiruri paralele.

Dacă perechile de valori corelate x_i, y_i sunt ponderate cu n_i, atunci sistemul de ecuaţii normale (15.50) devine:

$$\begin{cases} a\sum_{i=1}^{k} n_i + b\sum_{i=1}^{k} x_i n_i = \sum_{i=1}^{k} \frac{n_i}{y_i} \\ a\sum_{i=1}^{k} x_i n_i + b\sum_{i=1}^{k} x_i^2 n_i = \sum_{i=1}^{k} \frac{x_i n_i}{y_i} \end{cases} \tag{15.53}$$

Scanned with OKEN Scanner

$$a = \frac{\sum_{i=1}^{k} x_i^2 n_i \sum_{i=1}^{k} \frac{n_i}{y_i} - \sum_{i=1}^{k} \frac{x_i n_i}{y_i} \sum_{i=1}^{k} x_i n_i}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2} \tag{15.54}$$

$$b = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} \frac{x_i n_i}{y_i} - \sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} \frac{n_i}{y_i}}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2} \tag{15.55}$$

În cazul prezentării variabilelor empirice sub forma unui tabel de corelaţie cu dublă intrare, sistemul de ecuaţii normale necesar pentru determinarea parametrilor ecuaţiei de regresie este de forma:

$$\begin{cases} a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} x_i n_{i.} = \sum_{j=1}^{k} \frac{n_{.j}}{y_j} \\ a\sum_{i=1}^{s} x_i n_{i.} + b\sum_{i=1}^{s} x_i^2 n_{i.} = \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{x_i}{y_j} n_{ij} \end{cases} \tag{15.56}$$

de unde

$$a = \frac{\sum_{i=1}^{s} x_i^2 n_{i.} \sum_{j=1}^{k} \frac{n_{.j}}{y_j} - \sum_{i=1}^{s} x_i n_{i.} \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{x_i}{y_j} n_{ij}}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} x_i^2 n_{i.} - \left(\sum_{i=1}^{s} x_i n_{i.}\right)^2} \tag{15.57}$$

$$b = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} \frac{x_i}{y_j} n_{ij} - \sum_{i=1}^{s} x_i n_{i.} \sum_{j=1}^{k} \frac{n_{.j}}{y_j}}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} x_i^2 n_{i.} - \left(\sum_{i=1}^{s} x_i n_{i.}\right)^2} \tag{15.58}$$

Scanned with OKEN Scanner

Legătura hiperbolică, de tipul (15.36), se liniarizează prin substituţia $y = \frac{1}{y'}$, în ecuaţia $\frac{1}{y'} = a + bx$ şi se obţine $y = a + bx$

Expresiile coeficientului de corelaţie, în cazul legăturii hiperbolice (15.36) liniarizate, în funcţie de modul de sistematizare a datelor empirice cu privire la cele două variabile corelate, se prezintă astfel:

- în primul caz:

$$r_{x,\frac{1}{y}} = \frac{n\sum_{i=1}^{n}\frac{x_i}{y_i} - \sum_{i=1}^{n}x_i\sum_{i=1}^{n}\frac{1}{y_i}}{\sqrt{\left[n\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2\right]\left[n\sum_{i=1}^{n}\frac{1}{y_i^2} - \left(\sum_{i=1}^{n}\frac{1}{y_i}\right)^2\right]}} \qquad (15.59)$$

- în al doilea caz:

$$r_{x,\frac{1}{y}} = \frac{\sum_{i=1}^{k}n_i\sum_{i=1}^{k}\frac{n_ix_i}{y_i} - \sum_{i=1}^{k}n_ix_i\sum_{i=1}^{k}\frac{n_i}{y_i}}{\sqrt{\left[\sum_{i=1}^{k}n_i\sum_{i=1}^{k}n_ix_i^2 - \left(\sum_{i=1}^{k}n_ix_i\right)^2\right]\left[\sum_{i=1}^{k}n_i\sum_{i=1}^{k}\frac{n_i}{y_i^2} - \left(\sum_{i=1}^{k}\frac{n_i}{y_i}\right)^2\right]}} \qquad (15.60)$$

- în al treilea caz:

$$r_{x,\frac{1}{y}} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}\sum_{j=1}^{k}\frac{n_{ij}x_{i.}}{y_j} - \sum_{i=1}^{s}n_{i.}x_i\sum_{j=1}^{k}\frac{n_{.j}}{y_j}}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}n_{i.}x_i^2 - \left(\sum_{i=1}^{s}n_{i.}x_i\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{j=1}^{k}\frac{n_{.j}}{y_j^2} - \left(\sum_{j=1}^{k}\frac{n_{.j}}{y_j}\right)^2\right]}} \qquad (15.61)$$

În condiţiile legăturii hiperbolice (15.37) dată de ecuaţia:

Scanned with OKEN Scanner

$$y' = \frac{x'}{a'x' + b} = \frac{1}{a' + \frac{b'}{x}}$$

substituind pe $y = \frac{1}{y'}$ şi pe $x = \frac{1}{x'}$, se obţine: $y = a' + b'x$

Coeficientul de corelaţie pentru această legătură hiperbolică liniarizată, în funcţie de modul de sistematizare a datelor iniţiale cu privire la variabilele corelate, se prezintă astfel:

- în primul caz:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{n\sum_{i=1}^{s}\frac{1}{x_i}\frac{1}{y_i} - \sum_{i=1}^{s}\frac{1}{x_i}\sum_{i=1}^{s}\frac{1}{y_i}}{\sqrt{\left[n\sum_{i=1}^{s}\frac{1}{x_i^2} - \left(\sum_{i=1}^{s}\frac{1}{x_i}\right)^2\right]\left[n\sum_{i=1}^{s}\frac{1}{y_i^2} - \left(\sum_{i=1}^{s}\frac{1}{y_i}\right)^2\right]}} \qquad (15.62)$$

- în al doilea caz:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i}{x_i y_i} - \sum_{i=1}^{k}\frac{n_i}{x_i}\sum_{i=1}^{k}\frac{n_i}{y_i}}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i}{x_i^2} - \left(\sum_{i=1}^{k}\frac{n_i}{x_i}\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i}{y_i^2} - \left(\sum_{i=1}^{k}\frac{n_i}{y_i}\right)^2\right]}} \qquad (15.63)$$

- în al treilea caz:

$$r_{\frac{1}{x},\frac{1}{y}} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k}\frac{1}{x_i}\frac{1}{y_j} n_{ij} - \sum_{i=1}^{s}\frac{n_{i.}}{x_i}\sum_{j=1}^{k}\frac{n_{.j}}{y_j}}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\frac{n_{i.}}{x_i^2} - \left(\sum_{i=1}^{s}\frac{n_{i.}}{x_i}\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k}\frac{n_{.j}}{y_j^2} - \left(\sum_{j=1}^{k}\frac{n_{.j}}{y_j}\right)^2\right]}} \qquad (15.64)$$

15.1.3. Legături după modelul unei funcţii putere

Regresia după modelul unei funcţii putere este descrisă de ecuaţia:

$$y = ax^b \quad \text{sau} \quad \lg y = \lg a + b \lg x \qquad (15.65)$$

Scanned with OKEN Scanner

Tabelul 15.8. Calculele preliminare pentru determinarea indicatorilor de regresie şi corelaţie

Nr. crt	X / Y / x_i / y_i		-4	4 - 8	8 - 12	12 - 16	16 - 20	20 şi peste	Total $n_{.j}$	$y_j n_{.j}$	$y_j^2 n_{.j}$	$\sum_{j=1}^{k} \frac{y_j}{x_i} n_{ij}$	$\bar{x}/y_i$
			2	6	10	14	18						
0	A		2	3	4	5	6	7	8	9	10	11	12
1	- 3,6	3,3	-	-	-	-	4	5	9	29,7	98,01	1,4833	20,2
2	3,6 - 4,2	3,9	-	-	5	3	1	2	11	42,9	167,31	3,3569	14,0
3	4,2 - 4,8	4,5	-	5	3	-	-	-	8	36,0	162,00	5,1000	7,5
4	4,8 - 5,4	5,1	1	4	-	-	-	-	5	25,5	130,05	5,9500	5,2
5	5,4 - 6,0	5,7	3	2	-	-	-	-	5	28,5	162,45	10,4500	3,6
6	6,0 +	6,3	6	-	-	-	-	-	6	37,8	238,14	18,9000	2,0
7	Total $n_{i.}$		10	11	8	3	5	7	44	200,4	957,96	45,2403	-
8	$\frac{n_{i.}}{x_i}$		5,0000	1,8333	0,8000	0,2143	0,2778	0,3182	8,4436				
9	$\frac{n_{i.}}{x_i^2}$		2,5000	0,3056	0,0800	0,0153	0,0154	0,0145	2,9308				
10	$\sum_{i=1}^{s} \frac{y_j}{x_i} n_{ij}$		30,0000	9,0500	3,3000	0,8357	0,9500	1,1045	45,2403				
11	y_x		6,15	4,42	4,08	3,93	3,85	3,80	-				
12	$\bar{y}/x_i$		6,0	4,94	4,13	3,9	3,42	3,47	-				

Scanned with OKEN Scanner

Expresiile geometrice descrise de ecuaţia (15.65) se prezintă în figura 15.9.

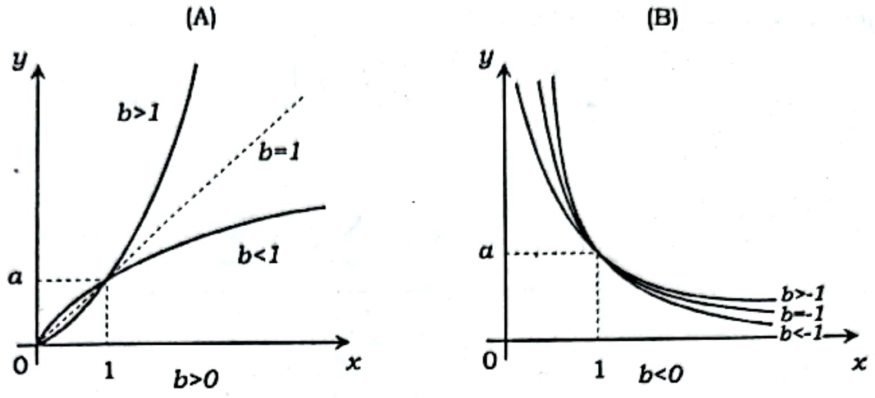


Figura 15.9. Funcţia de regresie tip putere $y = ax^b$

Observăm din figura 15.9, că funcţia tip putere trece prin punctul de coordonate $(1, a)$, oricare ar fi valorile parametrilor a şi b. În cazul în care coeficientul de regresie b este pozitiv (fig. 15.9.A) şi cuprins între 0 şi 1, curba descrisă de funcţia putere trece prin originea axelor de coordonate, este crescătoare şi convexă când $b<1$, iar pentru $b>1$ curba este tot crescătoare dar concavă şi când $b=1$ curba se transformă într-o dreaptă ce trece prin origine. Când b este negativ, curbele trec prin punctul de coordonate $(1, a)$ şi au ca asimptote axele de coordonate (fig. 15.9 B).

În cazul funcţiilor putere, legătura dintre variabile se manifestă astfel : unor creşteri relative (procentuale), egale ale variabilei X le corespund creşteri relative (procentuale) egale pentru variabila Y.

Sistemul de ecuaţii normale, obţinut pe baza metodei celor mai mici patrate, în cazul unei funcţii putere este de forma:

$$\begin{cases} n \lg a + b\sum_{i=1}^{n} \lg x_i = \sum_{i=1}^{n} \lg y_i \\ \lg a \sum_{i=1}^{n} \lg x_i + b\sum_{i=1}^{n} (\lg x_i)^2 = \sum_{i=1}^{n} (\lg x_i)(\lg y_i) \end{cases} \qquad (15.66)$$

de unde rezultă:

$$\lg a = \frac{\sum_{i=1}^{n} \lg y_i \sum_{i=1}^{n} (\lg x_i)^2 - \sum_{i=1}^{n} (\lg x_i) \sum_{i=1}^{n} (\lg x_i)(\lg y_i)}{n \sum_{i=1}^{n} (\lg x_i)^2 - \left(\sum_{i=1}^{n} \lg x_i \right)^2} \tag{15.67}$$

$$b = \frac{n \sum_{i=1}^{n} (\lg x_i)(\lg y_i) - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} \lg y_i}{n \sum_{i=1}^{n} (\lg x_i)^2 - \left(\sum_{i=1}^{n} \lg x_i \right)^2} \tag{15.68}$$

Dacă perechile de valori empirice corelate x_i, y_i sunt ponderate cu frecvenţa n_i, atunci sistemul de ecuaţii normale (15.66) devine:

$$\begin{cases} \lg a \sum_{i=1}^{k} n_i + b \sum_{i=1}^{k} n_i \lg x_i = \sum_{i=1}^{k} n_i \lg y_i \\ \lg a \sum_{i=1}^{k} n_i \lg x_i + b \sum_{i=1}^{k} n_i (\lg x_i)^2 = \sum_{i=1}^{k} n_i (\lg x_i)(\lg y_i) \end{cases} \tag{15.69}$$

Prin rezolvarea sistemului (15.69), rezultă:

$$\lg a = \frac{\sum_{i=1}^{k} n_i \lg y_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \sum_{i=1}^{k} n_i \lg x_i \sum_{i=1}^{k} n_i (\lg x_i)(\lg y_i)}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \left(\sum_{i=1}^{k} n_i \lg x_i \right)^2} \tag{15.70}$$

$$b = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)(\lg y_i) - \sum_{i=1}^{k} n_i \lg x_i \sum_{i=1}^{k} n_i \lg y_i}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \left(\sum_{i=1}^{k} n_i \lg x_i \right)^2} \tag{15.71}$$

Scanned with OKEN Scanner

În cazul prezentării datelor iniţiale, cu privire la variaţia fenomenelor corelate, sub formă de tabel de corelaţie cu dublă intrare, sistemul de ecuaţii normale va fi de forma:

$$\begin{cases} \lg a \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} n_{i.} \lg x_i = \sum_{j=1}^{k} n_{.j} \lg y_j \\ \lg a \sum_{i=1}^{s} n_{i.} \lg x_i + b\sum_{i=1}^{s} n_{i.} (\lg x_i)^2 = \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} (\lg x_i)(\lg y_j) \end{cases} \qquad (15.72)$$

Din rezolvarea sistemului (15.72), rezultă formulele uzuale de calcul al parametrilor ecuaţiei de regresie de tip funcţie putere (15.65):

$$\lg a = \frac{\sum_{j=1}^{k} n_{.j} \lg y_j \sum_{i=1}^{s} n_{i.} (\lg x_i)^2 - \sum_{i=1}^{s} n_{i.} \lg x_i \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} (\lg x_i)(\lg y_j)}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} (\lg x_i)^2 - \left(\sum_{i=1}^{s} n_{i.} \lg x_i\right)^2} \qquad (15.73)$$

$$b = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} (\lg x_i)(\lg y_j) - \sum_{i=1}^{s} n_{i.} \lg x_i \sum_{j=1}^{k} n_{.j} \lg y_j}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} (\lg x_i)^2 - \left(\sum_{i=1}^{s} n_{i.} \lg x_i\right)^2} \qquad (15.74)$$

Liniarizarea funcţiei putere:

$$y = ax^b$$

se obţine pe baza substituţiei $y = \lg y'$ şi $x = \lg x'$ în ecuaţia funcţiei putere.

Coeficientul de corelaţie r, calculat pe baza funcţiei putere liniarizate, în cele trei cazuri posibile de sistematizare a valorilor variabilelor empirice, se prezintă astfel:

- în primul caz:

Scanned with OKEN Scanner

$$r_{\lg x,\lg y}=\frac{n\sum_{i=1}^{n}(\lg x_i)(\lg y_i)-\sum_{i=1}^{n}\lg x_i\sum_{i=1}^{n}\lg y_i}{\sqrt{\left[n\sum_{i=1}^{n}(\lg x_i)^2-\left(\sum_{i=1}^{n}\lg x_i\right)^2\right]\left[n\sum_{i=1}^{n}(\lg y_i)^2-\left(\sum_{i=1}^{n}\lg y_i\right)^2\right]}} \tag{15.75}$$

- în al doilea caz:

$$r_{\lg x,\lg y}=\frac{\sum_{i=1}^{k}n_i\sum_{i=1}^{k}n_i(\lg x_i)(\lg y_i)-\sum_{i=1}^{k}n_i\lg x_i\sum_{i=1}^{k}n_i\lg y_i}{\sqrt{\left[\sum_{i=1}^{k}n_i\sum_{i=1}^{k}n_i(\lg x_i)^2-\left(\sum_{i=1}^{k}n_i\lg x_i\right)^2\right]\left[\sum_{i=1}^{k}n_i\sum_{i=1}^{k}n_i(\lg y_i)^2-\left(\sum_{i=1}^{k}n_i\lg y_i\right)^2\right]}} \tag{15.76}$$

- în al treilea caz:

$$r_{\lg x,\lg y}=\frac{\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}(\lg x_i)(\lg y_j)-\sum_{i=1}^{s}n_{i\cdot}\lg x_i\sum_{j=1}^{k}n_{\cdot j}\lg y_j}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}n_{i\cdot}(\lg x_i)^2-\left(\sum_{i=1}^{s}n_{i\cdot}\lg x_i\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{j=1}^{k}n_{\cdot j}(\lg y_j)^2-\left(\sum_{j=1}^{k}n_{\cdot j}\lg y_j\right)^2\right]}} \tag{15.77}$$

15.1.4. Legături de tip exponenţial

Funcţia exponenţială de tip clasic este de forma:

$$y=ae^{bx}\quad(-\infty<x<+\infty) \tag{15.78}$$

Curba reprezentativă a funcţiei exponenţiale se prezintă în figura 15.10.

Scanned with OKEN Scanner

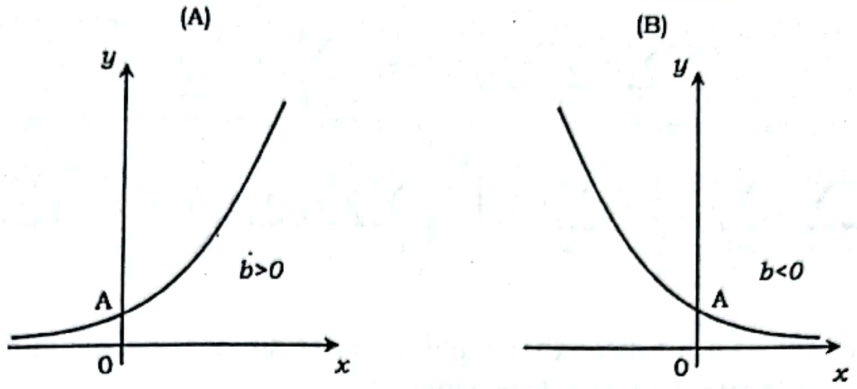


Figura 15.10. Funcţia exponenţială $y = ae^{bx}$ 1

Din figura 15.10 se observă că toate curbele exponenţiale trec prin punctul de coordonate $(0, a)$, iar axa absciselor este o asimptotă orizontală. Dacă coeficientul de regresie b este pozitiv, atunci curba exponenţială este crescătoare (fig. 15.10 (A)), iar dacă b este negativ, curba exponenţială este descrescătoare (fig. 15.10 (B)).

Liniarizarea legăturii (15.78) se obţine prin substituţia $y_i = ln\, y_i$ şi $a_i = ln\, a_i$.

Intensitatea legăturii dintre variabile se poate acum determina cu ajutorul coeficientului de corelaţie ce se calculează după una din formulele:

• în cazul perechilor x_i, y_i neponderate:

$$r_{x,ln\,y} = \frac{n\sum_{i=1}^{n} x_i \ln y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \ln y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} (\ln y_i)^2 - \left(\sum_{i=1}^{n} \ln y_i\right)^2\right]}} \qquad (15.79)$$

• în cazul perechilor de valori corelate (x_i, y_i) ponderate cu n_i:

$$r_{x,ln\,y} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i x_i \ln y_i - \sum_{i=1}^{k} x_i n_i \sum_{i=1}^{k} n_i \ln y_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} x_i^2 n_i - \left(\sum_{i=1}^{k} x_i n_i\right)^2\right]\left[\sum_{i=1}^{k} n_i (\ln y_i)^2 - \left(\sum_{i=1}^{k} n_i \ln y_i\right)^2\right]}} \qquad (15.80)$$

• în cazul unui tabel de corelaţie:

$$r_{x,\ln y}=\frac{\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}x_i\ln y_j-\sum_{i=1}^{s}x_i n_{i.}\sum_{j=1}^{k}(\ln y_j)n_{.j}}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{i=1}^{s}n_{i.}x_i^2-\left(\sum_{i=1}^{s}x_i n_{i.}\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}\sum_{j=1}^{k}n_{.j}(\ln y_j)^2-\left(\sum_{j=1}^{k}n_{.j}\ln y_j\right)^2\right]}} \tag{15.81}$$

O altă formă a funcţiei exponenţiale, care se situează numai în cadranul întâi al axelor de coordonate, este dată de ecuaţia:

$$y=ae^{\frac{b}{x}} \qquad (0<x<\infty) \tag{15.82}$$

Curba reprezentativă a funcţiei (15.82) se reprezintă în figura 15.11.

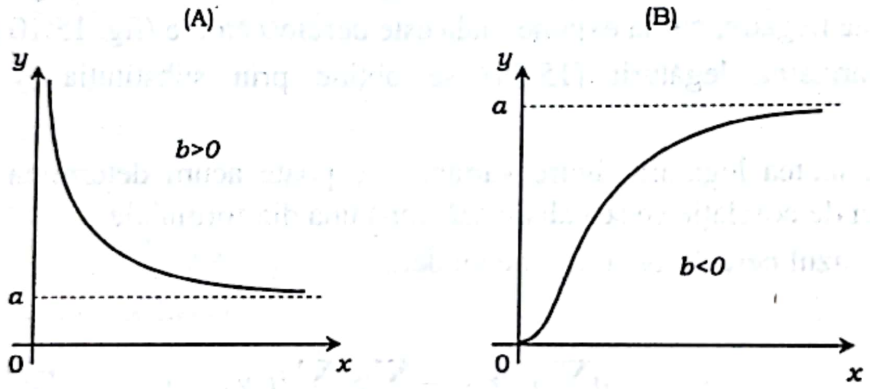


Figura 15.11. Funcţia exponenţială $y=ae^{\frac{b}{x}}$

Dacă coeficientul de regresie b este pozitiv, curba exponenţială este descrescătoare (fig. 15.11 (A)), iar dacă b este negativ, curba exponenţială este crescătoare (fig. 15.11 (B)).

Curba exponenţială (A) are două asimptote: una orizontală $y=a$ şi alta verticală $x=0$.

Curba exponenţială (B) se caracterizează prin următoarele trăsături:

- trece prin originea axelor de coordonate;
- are o asimptotă orizontală $y=a$;
- are un punct de inflexiune de coordonate $(-\frac{b}{a},\frac{e}{a^2})$.

Scanned with OKEN Scanner

Liniarizarea legăturii exponenţiale dintre variabilele (15.12) se obţine prin efectuarea următoarelor substituţii: $y_i' = ln\, y_i$ şi $x_i' = \frac{1}{x_i}$ în ecuaţia $y' = a' e^{bx}$.

Intensitatea legăturii dintre variabile, sub formă liniarizată, se poate calcula cu ajutorul coeficientului de corelaţie liniară între $\frac{1}{x_i}$ şi $ln\, y_i$. Corespunzător celor trei modalităţi de prezentare a datelor statistice, coeficientul de corelaţie va lua următoarele expresii:

- în primul caz:

$$r_{\frac{1}{x},ln\,y} = \frac{n\sum_{i=1}^{n}\frac{ln\,y_i}{x_i} - \sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n} ln\,y_i}{\sqrt{\left[n\sum_{i=1}^{n}\frac{1}{x_i^2} - \left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2\right]\left[n\sum_{i=1}^{n}(ln\,y_i)^2 - \left(\sum_{i=1}^{n} ln\,y_i\right)^2\right]}} \qquad (15.83)$$

- în al doilea caz:

$$r_{\frac{1}{x},ln\,y} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i\, ln\,y_i}{x_i} - \sum_{i=1}^{k}\frac{n_i}{x_i}\sum_{i=1}^{k} n_i\, ln\,y_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i}{x_i^2} - \left(\sum_{i=1}^{k}\frac{n_i}{x_i}\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (ln\,y_i)^2 - \left(\sum_{i=1}^{k} n_i\, ln\,y_i\right)^2\right]}} \qquad (15.84)$$

- în al treilea caz:

$$r_{\frac{1}{x},ln\,y} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k}\frac{ln\,y_j}{x_i} n_{ij} - \sum_{i=1}^{s}\frac{n_{i.}}{x_i}\sum_{j=1}^{k} n_{.j}\, ln\,y_j}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\frac{n_{i.}}{x_i^2} - \left(\sum_{i=1}^{s}\frac{n_{i.}}{x_i}\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k} n_{.j} (ln\,y_j)^2 - \left(\sum_{j=1}^{k} n_{.j}\, ln\,y_j\right)^2\right]}} \qquad (15.85)$$

O altă formă modificată a funcţiei exponenţiale este următoarea:

Scanned with OKEN Scanner

$$y = \frac{1}{a + be^{-x}} \qquad (-\infty < x < +\infty) \qquad (15.86)$$

Curba exponenţială descrisă de funcţia (15.86) se prezintă în figura 15.12.

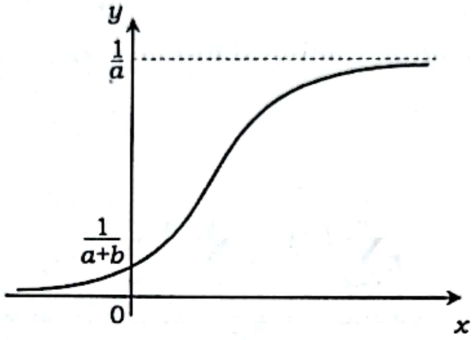


Figura 15.12. Curba exponenţială $y = \frac{1}{a + be^{-x}}$

Se observă, din figura 15.12., că această curbă exponenţială are două asimptote orizontale, una este chiar axa absciselor $y = 0$ şi cealaltă, paralelă cu cea dintâi, este $y = \frac{1}{a}$. Curba exponenţială intersectează axa ordonatelor în punctul de coordonate $(0, \frac{1}{a+b})$ şi are un punct de inflexiune $(ln\frac{b}{a}, \frac{1}{2a})$.

Mai des utilizate în practică sunt două forme particulare ale acestei curbe şi anume, curba lui *Gompertz* şi curba logistică de care ne vom ocupa în capitolul referitor la analiza seriilor dinamice.

Liniarizarea legăturii exponenţiale are loc dacă în formula (15.86) se fac substituţile: $y = \frac{1}{y_i'}$ şi $x' = e^{-x}$.

Intensitatea legăturii, în acest tip de funcţie exponenţială, după liniarizare, se poate determina cu ajutorul uneia din formule, în baza modului de prezentare a datelor empirice:

- când perechile de valori x_i, y_i sunt neponderate:

Scanned with OKEN Scanner

$$r_{e^{-x},\frac{1}{y}} = \frac{n\sum_{i=1}^{n} e^{-x_i}\frac{1}{y_i} - \sum_{i=1}^{n} e^{-x_i}\sum_{i=1}^{n}\frac{1}{y_i}}{\sqrt{\left[n\sum_{i=1}^{n}(e^{-x_i})^2 - \left(\sum_{i=1}^{n} e^{-x_i}\right)^2\right]\left[n\sum_{i=1}^{n}\frac{1}{y_i^2} - \left(\sum_{i=1}^{n}\frac{1}{y_i}\right)^2\right]}} \tag{15.87}$$

- când perechile de valori x_i, y_i sunt ponderate cu n_i:

$$r_{e^{-x},\frac{1}{y}} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i e^{-x_i}\frac{1}{y_i} - \sum_{i=1}^{k} n_i e^{-x_i}\sum_{i=1}^{k}\frac{1}{y_i}}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (e^{-x_i})^2 - \left(\sum_{i=1}^{k} n_i e^{x_i}\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k}\frac{n_i}{y_i^2} - \left(\sum_{i=1}^{k}\frac{n_i}{y_i}\right)^2\right]}} \tag{15.88}$$

- când variabilele X şi Y se prezintă sub forma unui tabel de corelaţie:

$$r_{e^{-x},\frac{1}{y}} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} e^{-x_i}\frac{1}{y_j} - \sum_{i=1}^{s} n_{i.} e^{-x_i}\sum_{j=1}^{k}\frac{n_{.j}}{y_j}}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.}(e^{-x_i})^2 - \left(\sum_{i=1}^{s} n_{i.} e^{-x_i}\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k}\frac{n_{.j}}{y_j^2} - \left(\sum_{j=1}^{k}\frac{n_{.j}}{y_j}\right)^2\right]}} \tag{15.89}$$

O formă transformată a ecuaţiei exponenţiale, convenabilă pentru aplicaţiile practice, este:

$$y = ab^x \tag{15.90}$$

sau

$$\lg y = \lg a + x \lg b$$

Pentru determinarea parametrilor ecuaţiei de regresie (15.90), prin aplicarea metodei celor mai mici pătrate, se obţine sistemul de ecuaţii normale:

Scanned with OKEN Scanner

$$\begin{cases} n\lg a+\lg b\sum_{i=1}^{s}x_i=\sum_{i=1}^{s}\lg y_i \\ \lg a\sum_{i=1}^{s}x_i+\lg b\sum_{i=1}^{s}x_i^2=\sum_{i=1}^{s}x_i\lg y_i \end{cases} \quad (15.91)$$

Prin rezolvarea sistemului (15.91) se obţin formulele uzuale ale parametrilor ecuaţiei de regresie.

Sistemul (15.91) este aplicabil în primul caz de sistematizare a datelor cu privire la cele două variabile. În celelalte două cazuri, sistemul ia forma:

- în cazul al doilea:

$$\begin{cases} \lg a\sum_{i=1}^{k}n_i+\lg b\sum_{i=1}^{k}x_i^2n_i=\sum_{i=1}^{k}n_i\lg y_i \\ \lg a\sum_{i=1}^{k}x_in_i+\lg b\sum_{i=1}^{k}x_i^2n_i=\sum_{i=1}^{k}n_ix_i\lg y_i \end{cases} \quad (15.92)$$

- în cazul al treilea:

$$\begin{cases} \lg a\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}+\lg b\sum_{i=1}^{s}x_in_{i.}=\sum_{j=1}^{k}n_{.j}\lg y_j \\ \lg a\sum_{i=1}^{s}x_in_{i.}+\lg b\sum_{i=1}^{s}x_i^2n_{i.}=\sum_{i=1}^{s}\sum_{j=1}^{k}n_{ij}x_i\lg y_j \end{cases} \quad (15.93)$$

Intensitatea corelaţiei exponenţiale (15.90) se determină cu ajutorul raportului de corelaţie. Raportul de corelaţie, în funcţie de modul de prezentare a datelor iniţiale, se determină astfel:

- în primul caz:

Scanned with OKEN Scanner

$$\eta = \sqrt{\frac{a\sum_{i=1}^{n} lg\, y_i + lg\, b\sum_{i=1}^{n} x_i\, lg\, y_i - \frac{\left(\sum_{i=1}^{n} lg\, y_i\right)^2}{n}}{\sum_{i=1}^{n} (lg\, y_i)^2 - \frac{\left(\sum_{i=1}^{n} lg\, y_i\right)^2}{n}}} \qquad (15.94)$$

- în cazul al doilea:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} n_i\, lg\, y_i + lg\, b\sum_{i=1}^{k} n_i x_i\, lg\, y_i - \frac{\left(\sum_{i=1}^{k} n_i\, lg\, y_i\right)^2}{\sum_{i=1}^{k} n_i}}{\sum_{i=1}^{k} n_i (lg\, y_i)^2 - \frac{\left(\sum_{i=1}^{k} n_i\, lg\, y_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.95)$$

- în cazul al treilea:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} n_{.j}\, lg\, y_j + lg\, b\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i\, lg\, y_j - \frac{\left(\sum_{j=1}^{k} n_{.j}\, lg\, y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{j=1}^{k} n_{.j} (lg\, y_j)^2 - \frac{\left(\sum_{j=1}^{k} n_{.j}\, lg\, y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.96)$$

Scanned with OKEN Scanner

Dacă în ecuaţia (15.90) se face substituţia $lg\, y' = y$, se liniarizează legătura exponenţială. Intensitatea legăturii se poate măsura cu ajutorul coeficientului de corelaţie liniară dintre x şi $\lg y$. Corespunzător celor trei cazuri de sistematizare a datelor iniţiale, formula coeficientului de corelaţie devine:

- în primul caz:

$$r_{x,\lg y} = \frac{n\sum_{i=1}^{n} x_i \lg y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} \lg y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} (\lg y_i)^2 - \left(\sum_{i=1}^{n} \lg y_i\right)^2\right]}} \qquad (15.97)$$

- în cazul al doilea:

$$r_{x,lg\,y} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i x_i \, lg\, y_i - \sum_{i=1}^{k} n_i x_i \sum_{i=1}^{k} n_i \, lg\, y_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i x_i^2 - \left(\sum_{i=1}^{k} n_i x_i\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (lg\, y_i)^2 - \left(\sum_{i=1}^{k} n_i \, lg\, y_i\right)^2\right]}} \qquad (15.98)$$

- în cazul al treilea:

$$r_{x,lg\,y} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} x_i \, lg\, y_j - \sum_{i=1}^{s} n_{i.} x_i \sum_{j=1}^{k} n_{.j} \, lg\, y_i}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} x_i^2 - \left(\sum_{i=1}^{s} n_{i.} x_i\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k} n_{.j} (lg\, y_j)^2 - \left(\sum_{j=1}^{k} n_{.j} \, lg\, y_j\right)^2\right]}} \qquad (15.99)$$

15.1.5. Legături de tip logaritmic

Funcţia logaritmică (fig.15.13) este dată de expresia:

$$y = a + b \, lg\, x \qquad (0 < x < \infty) \qquad (15.100)$$

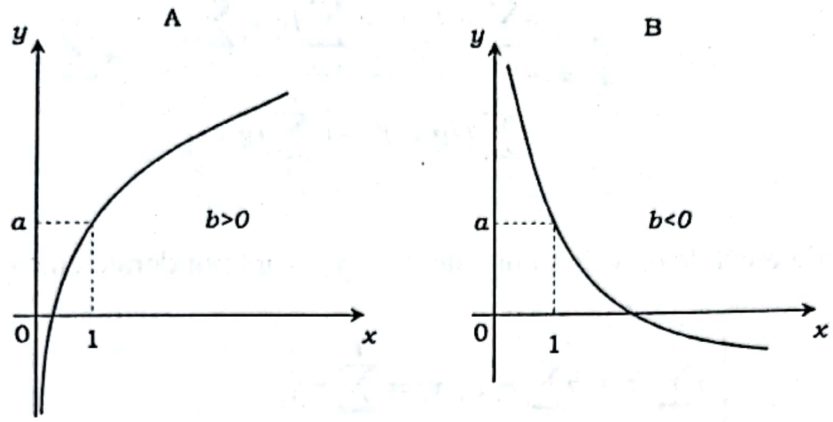


Figura 15.13. Funcţia logaritmică $y = a + b \lg x$

Din figura 15.13. se observă că, dacă coeficientul de regresie b este pozitiv, curba este crescătoare (fig.15.13 (A)), iar dacă b este negativ, curba este descrescătoare (fig. 15.13 (B)). De asemenea, ambele curbe trec prin punctul de coordonate $(1, a)$ şi au ca asimptotă verticală axa coordonatelor $x = 0$.

Parametrii funcţiei de regresie (15.100) se pot determina cu ajutorul metodei celor mai mici pătrate. Prin aplicarea acestei metode se obţin, în funcţie de modul de prezentare a celor două variabile corelate, următoarele sisteme de ecuaţii normale:

- când perechile de valori (x_i, y_i) sunt neponderate:

$$\begin{cases} na + b\sum_{i=1}^{n} \lg x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} \lg x_i + b\sum_{i=1}^{n} (\lg x_i)^2 = \sum_{i=1}^{n} y_i \lg x_i \end{cases} \qquad (15.101)$$

iar

$$a = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} (\lg x_i)^2 - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} y_i \lg x_i}{n\sum_{i=1}^{n} (\lg x_i)^2 - \left(\sum_{i=1}^{n} \lg x_i\right)^2} \qquad (15.102)$$

$$b = \frac{n\sum_{i=1}^{n} y_i \lg x_i - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} (\lg x_i)^2 - \left(\sum_{i=1}^{n} \lg x_i\right)^2} \tag{15.103}$$

• când perechile de valori corelate (x_i, y_i) sunt ponderate cu frecvenţa n_i:

$$\begin{cases} a\sum_{i=1}^{k} n_i + b\sum_{i=1}^{k} n_i \lg x_i = \sum_{i=1}^{k} n_i y_i \\ a\sum_{i=1}^{k} n_i \lg x_i + b\sum_{i=1}^{k} n_i (\lg x_i)^2 = \sum_{i=1}^{k} n_i y_i \lg x_i \end{cases} \tag{15.104}$$

de unde rezultă:

$$a = \frac{\sum_{i=1}^{k} n_i y_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \sum_{i=1}^{k} n_i \lg x_i \sum_{i=1}^{k} n_i y_i \lg x_i}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \left(\sum_{i=1}^{k} n_i \lg x_i\right)^2} \tag{15.105}$$

$$b = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i y_i \lg x_i - \sum_{i=1}^{k} n_i \lg x_i \sum_{i=1}^{k} n_i y_i}{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \left(\sum_{i=1}^{k} n_i \lg x_i\right)^2} \tag{15.106}$$

• când cele două variabile corelate se prezintă sub forma unui tabel cu dublă intrare:

$$\begin{cases} a\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} + b\sum_{i=1}^{s} n_{i.} \lg x_i = \sum_{j=1}^{k} n_{.j} y_j \\ a\sum_{i=1}^{s} n_{i.} \lg x_i + b\sum_{i=1}^{s} n_{i.} (\lg x_i)^2 = \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} y_j \lg x_i \end{cases} \tag{15.107}$$

Scanned with OKEN Scanner

iar

$$a=\frac{\sum_{j=1}^{k} n_{.j} y_j \sum_{i=1}^{s} n_{i.} (lg\, x_i)^2 - \sum_{i=1}^{s} n_{i.} lg\, x_i \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} y_j lg\, x_i}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} (lg\, x_i)^2 - \left(\sum_{i=1}^{s} n_{i.} lg\, x_i\right)^2} \qquad (15.108)$$

$$b=\frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} y_j lg\, x_i - \sum_{i=1}^{s} n_{i.} lg\, x_i \sum_{j=1}^{k} n_{.j} y_j}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} (lg\, x_i)^2 - \left(\sum_{i=1}^{s} n_{i.} lg\, x_i\right)^2} \qquad (15.109)$$

Intensitatea legăturii logaritmice de tipul (15.101) se măsoară cu ajutorul raportului de corelaţie:

$$\eta=\sqrt{1-\frac{\sigma^2_{yy_x}}{\sigma^2_y}} \qquad \text{în care} \qquad y_x = a+b\, lg\, x \qquad (15.110)$$

Calculul simplificat al raportului de corelaţie în funcţie de forma de prezentare a celor două variabile corelate, se efectuează după una din următoarele formule:

- în primul caz:

$$\eta=\sqrt{\frac{a\sum_{i=1}^{n} y_i + b\sum_{i=1}^{n} y_i lg\, x_i - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}{\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}}} \qquad (15.111)$$

- în al doilea caz:

$$\eta = \sqrt{\frac{a\sum_{i=1}^{k} n_i y_i + b\sum_{i=1}^{k} n_i y_i \lg x_i - \frac{\left(\sum_{i=1}^{k} n_i y_i\right)^2}{\sum_{i=1}^{s} n_i}}{\sum_{i=1}^{k} n_i y_i^2 - \frac{\left(\sum_{i=1}^{k} n_i y_i\right)^2}{\sum_{i=1}^{k} n_i}}} \qquad (15.112)$$

- în al treilea caz:

$$\eta = \sqrt{\frac{a\sum_{j=1}^{k} n_{.j} y_j + b\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} y_j \lg x_i - \frac{\left(\sum_{j=1}^{k} n_{.j} y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}\sum_{j=1}^{k} n_{.j} y_j^2 - \frac{\left(\sum_{j=1}^{k} n_{.j} y_j\right)^2}{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij}}}} \qquad (15.113)$$

Ecuaţia logaritmică (15.100) se poate liniariza prin substituţiile:

$$x = \lg x' \qquad \text{şi} \qquad y = y'$$

$$y' = a + b \lg x' \rightarrow y = a + bx$$

Coeficientul de corelaţie dintre $\log x$ şi y va avea următoarele expresii:

- în cazul (x_i, y_i) neponderate:

$$r_{lg\,x,y} = \frac{n\sum_{i=1}^{n} y_i \lg x_i - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} (\lg x_i)^2 - \left(\sum_{i=1}^{n} \lg x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \qquad (15.114)$$

- în cazul al doilea:

$$r_{log\,x,y} = \frac{\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i y_i \lg x_i - \sum_{i=1}^{k} n_i \lg x_i \sum_{i=1}^{k} n_i y_i}{\sqrt{\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i (\lg x_i)^2 - \left(\sum_{i=1}^{k} n_i \lg x_i\right)^2\right]\left[\sum_{i=1}^{k} n_i \sum_{i=1}^{k} n_i y_i^2 - \left(\sum_{i=1}^{k} n_i y_i\right)^2\right]}} \qquad (15.115)$$

- în cazul al treilea:

$$r_{log\,x,y} = \frac{\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} y_j \lg x_i - \sum_{i=1}^{s} n_{i.} \lg x_i \sum_{j=1}^{k} n_{.j} y_i}{\sqrt{\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{i=1}^{s} n_{i.} (\lg x_i^2) - \left(\sum_{i=1}^{s} n_{i.} \lg x_i\right)^2\right]\left[\sum_{i=1}^{s}\sum_{j=1}^{k} n_{ij} \sum_{j=1}^{k} n_{.j} y_j^2 - \left(\sum_{j=1}^{k} n_{.j} y_j\right)^2\right]}} \qquad (15.116)$$

15.1.6. Alegerea celei mai potrivite ecuaţii de regresie

Una dintre problemele fundamentale ale regresiei şi corelaţiei este alegerea celui mai potrivit model de regresie. De modul cum este rezolvată această problemă depinde încrederea ce se poate acorda indicatorilor obţinuţi din calculele de regresie şi corelaţie.

Dacă există legităţi, pe care le cunoaştem apriori cu privire la forma legăturii dintre variabile, atunci vom utiliza drept ecuaţie de regresie expresia matematică a acestora. Dacă nu există o asemenea bază teoretică se procedează la o analiză calitativă a esenţei fenomenelor şi dacă se ajunge la concluzia că poate exista legătura

Scanned with OKEN Scanner

între variabile, pentru a alege cel mai potrivit model de regresie se pot utiliza următoarele reguli:

1. Se reprezintă grafic datele empirice şi se trasează vizual o curbă care să se apropie cât mai mult de punctele empirice şi apoi se alege o ecuaţie care să corespundă cât mai bine curbei empirice. Această metodă se poate utiliza în practică, numai dacă numărul perechilor de valori este suficient de mare $(n > 30)$.

2. Pentru alegerea în fiecare caz particular a curbei celei mai potrivite se poate utiliza şi metoda diferenţelor succesive ale variabilei Y.

• Când primele diferenţe ale variabilei X sunt egale atunci se recomandă utilizarea modelului liniar de regresie simplă: $y = a + bx$;

• Dacă diferenţele de ordinul al doilea sunt constante atunci se recomandă utilizarea modelului parabolei de gradul doi : $y = a + bx + cx^2$;

• Dacă diferenţele de ordinul al treilea sunt constante, atunci se recomandă utilizarea parabolei de gradul trei : $y = a + bx + c\,x^2 + d\,x^3$;

• În general, dacă diferenţele de ordinul n sunt constante, atunci se recomandă utilizarea polinomului de gradul n ;

• Dacă primele diferenţe ale logaritmilor valorilor lui Y sunt constante, atunci se recomandă utilizarea unei curbe semilogaritmice de forma: $lg\, y = a + bx$;

• Dacă primele diferenţe ale mărimilor inverse ale variabilei dependente Y sunt constante, se poate utiliza o curbă hiperbolică de forma: $\frac{1}{y} = a + bx$;

3. Dacă raportul dintre diferenţele de ordinul întâi ale variabilei X şi cele corespunzătoare pentru variabila dependentă y este constant se poate utiliza modelul unei curbe dublu logaritmice de forma:

$$lg\, y = a + b\, lg\, x$$

4. Se reprezintă datele empirice pe o hârtie milimetrică obişnuită:

• Dacă corelograma variabilelor X şi $\log Y$ ia o formă liniară, atunci legătura dintre variabile se poate aproxima printr-o curbă semilogaritmică:

$$lg\, y = a + bx$$

• Dacă corelograma logaritmilor variabilelor X şi Y ia o formă liniară, atunci legătura dintre variabile poate fi descrisă de o ecuaţie logaritmică;

Scanned with OKEN Scanner

• Dacă graficul perechilor de valori $(x, \frac{1}{y})$ ia forma unei drepte, atunci legătura dintre variabile este hiperbolică : $\frac{1}{y} = a + bx$

5. Dacă se reprezintă pe o hârtie semilogaritmică perechile de valori corelate (x_i, y_i)şi corelograma ia forma unei drepte, atunci legătura dintre variabile se apropie de curba semilogaritmică : $lg\, y = a + b\, lg\, x$

6. Dacă corelograma valorilor empirice ia forma unei drepte pe o reţea dublu logaritmică, atunci legătura dintre variabile poate fi descrisă de o funcţie putere:

$$lg\, y = lg\, a + b\, lg\, x$$

7. Dacă pe un grafic, în care pe abscisă scara este aritmetică, iar pe ordonată valorile reciproce, corelograma ia forma unei drepte, legătura dintre variabile se apropie de o hiperbolă descrisă de ecuaţia : $\frac{1}{y} = a + bx$

Desigur, în activitatea statistică diferenţele de diferite ordine se vor apropia mai mult sau mai puţin de egalitatea între ele şi numai în cazuri cu totul particulare ele vor fi chiar egale. Curbele empirice se vor apropia mai mult sau mai puţin de anumite modele teoretice şi acestea, chiar dacă nu ne vor spune dintr-o dată care este curba cea mai potrivită în fiecare caz în parte, ne vor ajuta totuşi să excludem din calcul anumite familii de curbe şi apoi, prin încercări succesive, se va aprecia care curbă este mai potrivită cu ajutorul diferitelor tehnici statistice.

Scanned with OKEN Scanner

16. REGRESIE ŞI CORELAŢIE MULTIPLĂ LINIARĂ

16.1. Noţiunea de legătură multiplă

Legăturile multiple de corelaţie se stabilesc între două sau mai multe caracteristici factoriale X_i şi o caracteristică rezultativă Y:

$$Y = f(X_i) + \varepsilon \qquad i = \overline{1,k} \tag{16.1}$$

unde ε este eroarea reziduală.

Astfel de legături se întâlnesc atât în cadrul fenomenelor naturale, cât şi în cazul fenomenelor sociale şi economice complexe.

Într-o legătură multiplă variabilele factoriale au influenţe diferite asupra variabilei rezultative, unele exercită o acţiune importantă asupra fenomenului efect şi trebuie să fie luate în calculele de regresie şi corelaţie, în timp ce altele au o acţiune mai puţin importantă şi pot fi neglijate. Modelele de corelaţie, ca oricare alte modele, simplifică realitatea, deoarece este foarte dificil să se ia în consideraţie mulţimea tuturor factorilor cauzali care acţionează asupra unui anumit fenomen. De aceea, factorii cauzali se ierarhizează în ordinea importanţei acţiunii lor asupra fenomenului efect şi se iau în calcul numai acei care au o acţiune mai importantă.

Ecuaţia (16.1) poate fi liniară sau curbilinie în funcţie de formele legăturilor dintre fiecare pereche de valori (Y, X_i). Dacă toate legăturile simple dintre Y şi X_i sunt liniare, atunci şi legătura multiplă de corelaţie este tot liniară, iar dacă cel puţin una dintre legăturile simple (Y, X_i)este curbilinie, atunci şi legătura multiplă este curbilinie.

Creşterea economică se poate exprima printr-o funcţie de regresie, după modelul funcţiilor de producţie de tip *Cobb-Douglas:*

$$Y = aX_1^{b_1} X_2^{b_2} + \varepsilon \tag{16.2}$$

în care: Y = venitul naţional;

X_1 = valoarea fondurilor fixe productive din cadrul economiei naţionale;

X_2 = forţa de muncă din sfera productivă.

Scanned with OKEN Scanner

Parametrii b_1 şi b_2 pot să reprezinte elasticităţi , iar suma lor este egală cu unitatea, dacă se consideră că Y depinde numai de X_1 şi X_2:

$$b_1 + b_2 = 1 \qquad de\ unde \qquad b_1 = 1 - b_2 \qquad si \qquad b_2 = 1 - b_1$$

Un alt economist *Edelberg* (1936) a introdus un al treilea factor - durata procesului de producţie X_3:

$$Y = aX_1^{b_1} \cdot X_2^{b_2} \cdot X_3^{b_3} + \varepsilon \tag{16.3}$$

în care $b_1 + b_2 + b_3 = 1$

Cercetările ulterioare au arătat că acest tip de funcţii se poate aplica atât la nivel micro, cât şi la nivel macroeconomic.

Remarcăm şi faptul că, pentru utilizarea în practică a acestor funcţii de regresie, se poate proceda la liniarizarea lor prin logaritmare.

16.2. Regresie şi corelaţie liniară multiplă

16.2.1. Regresie liniară multiplă

În cazul unei legături liniare multiple, ecuaţia (16.1) reprezintă o funcţie liniară de forma:

$$Y_{x_1,x_2,\ldots,x_k} = a_0 + a_1X_1 + a_2X_2 + \cdots + a_iX_i + \cdots + a_kX_k \tag{16.4}$$

în care:

a_0 - reprezintă parametrul care exprimă influenţa celorlalţi factori consideraţi cu acţiune constantă, în afară de cei K factori cauzali luaţi în calcul;

a_i - reprezintă coeficienţii de regresie multiplă care arată cu cât variază variabila rezultativă Y, atunci când variabila factorială X_i se modifică cu o unitate.

Legătura multiplă liniară se reprezintă grafic sub forma unei suprafeţe multidimensionale. În cazul corelaţiei dintre cele două variabile independente (X_1 şi X_2) şi o variabilă rezultativă Y, expresia geometrică se prezintă sub forma planului *ABCD* din figura 16.1.

Din modelul (16.4), se observă că, în cazul a k variabile factoriale, ecuaţia de regresie multiplă liniară depinde de $k+1$ parametri.

Scanned with OKEN Scanner

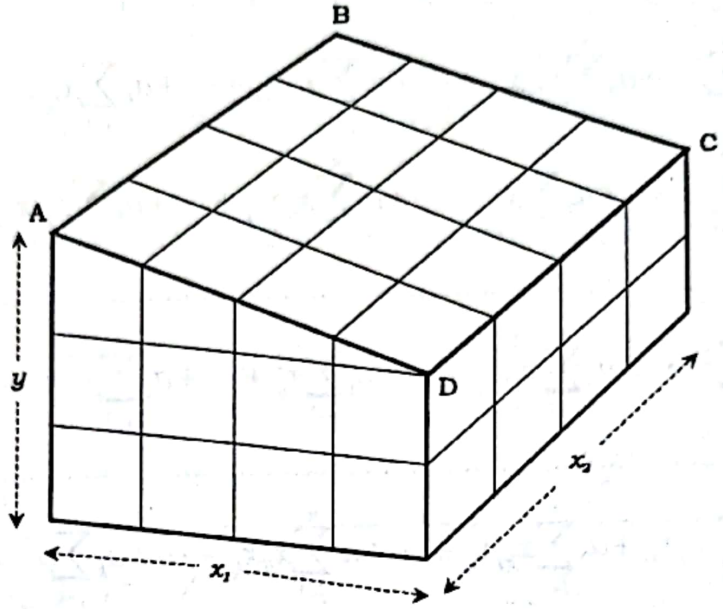


Figura 16.1. Legătură multiplă liniară dintre trei variabile

Parametrii $a_0, a_1, a_2, \ldots, a_i, \ldots, a_k$ se determină cu ajutorul metodei celor mai mici pătrate, punând condiţia ca suma pătratelor abaterilor dintre valorile empirice y_i şi valorile teoretice corespunzătoare $y_{x_1, x_2, \ldots, x_k}$ să fie minimă:

$$S = \sum_{j=1}^{n} (Y_j - Y_{x_1, x_2, \ldots, x_k})^2 = min\,im \qquad j = \overline{1, n} \tag{16.5}$$

sau

$$S = \sum_{j=1}^{n} (y_j - a_0 - a_1 x_{1j} - a_2 x_{2j} - \cdots - a_i x_{ij} - \cdots - a_k x_{kj})^2 = min\,im$$

Pentru a afla nivelul minim al funcţiei S se determină derivatele parţiale ale acestei funcţii în raport cu parametri $a_0, a_1, a_2, \ldots, a_i, \ldots, a_k$ care apoi se egalează cu zero şi se obţine:

Scanned with OKEN Scanner

$$na_0 + a_1\sum_{j=1}^{n}x_{1j} + a_2\sum_{j=1}^{n}x_{2j} + \cdots + a_i\sum_{j=1}^{n}x_{ij} + \cdots + a_k\sum_{j=1}^{n}x_{kj} = \sum_{j=1}^{n}y_j$$

$$a_0\sum_{j=1}^{n}x_{1j} + a_1\sum_{j=1}^{n}x_{1j}^2 + a_2\sum_{j=1}^{n}x_{2j}x_{1j} + \cdots + a_i\sum_{j=1}^{n}x_{ij}x_{1j} + \cdots + a_k\sum_{j=1}^{n}x_{kj}x_{1j} = \sum_{j=1}^{n}y_jx_{1j}$$

$$a_0\sum_{j=1}^{n}x_{2j} + a_1\sum_{j=1}^{n}x_{1j}x_{2j} + a_2\sum_{j=1}^{n}x_{2j}^2 + \cdots + a_i\sum_{j=1}^{n}x_{ij}x_{2j} + \cdots + a_k\sum_{j=1}^{n}x_{kj}x_{2j} = \sum_{j=1}^{n}y_jx_{2j}$$

.....

$$a_0\sum_{j=1}^{n}x_{ij} + a_1\sum_{j=1}^{n}x_{1j}x_{ij} + a_2\sum_{j=1}^{n}x_{2j}x_{ij} + \cdots + a_i\sum_{j=1}^{n}x_{ij}^2 + \cdots + a_k\sum_{j=1}^{n}x_{kj}x_{ij} = \sum_{j=1}^{n}y_jx_{ij}$$

.....

$$a_0\sum_{j=1}^{n}x_{kj} + a_1\sum_{j=1}^{n}x_{1j}x_{kj} + a_2\sum_{j=1}^{n}x_{2j}x_{kj} + \cdots + a_i\sum_{j=1}^{n}x_{ij}x_{kj} + \cdots + a_k\sum_{j=1}^{n}x_{kj}^2 = \sum_{j=1}^{n}y_jx_{kj} \qquad (16.6)$$

Dacă se împarte prima ecuaţie a sistemului (16.6) la n, rezultă:

$$a + a_1\bar{x}_1 + a_2\bar{x}_2 + \cdots + a_i\bar{x}_i + \cdots + a_k\bar{x}_k = \bar{y} \qquad (16.7)$$

Aceasta înseamnă că suprafaţa de regresie reprezintă un hiperplan în spaţiul euclidian cu $k+1$ dimensiuni şi trece prin punctul de coordonate $(\bar{y}, \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_i, \ldots, \bar{x}_k)$.

În cazul a două variabile independente şi una dependentă, sistemul de ecuaţii normale (16.6) se poate scrie substituind fiecare variabilă prin abaterile acesteia faţă de medie. Această transformare prezintă avantajul simplificării calculelor, deoarece se ştie că suma abaterilor unei variabile faţă de media lor este egală cu zero, adică:

$$\sum_{j=1}^{n}(x_{1j} - \bar{x}_1) = \sum_{j=1}^{n}(x_{2j} - \bar{x}_2) = \sum_{j=1}^{n}(y_j - \bar{y}) = 0$$

În aceste condiţii, prima ecuaţie a sistemului (16.6) se elimină şi acesta devine:

Scanned with OKEN Scanner

$$\begin{cases} a_1\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2 + a_2\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) = \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(y_j-\bar{y}) \\ a_1\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) + a_2\sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 = \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)(y_j-\bar{y}) \end{cases} \quad (16.8)$$

Rezolvând sistemul (16.8) prin metoda determinanţilor, rezultă:

$$a_1 = \frac{\begin{vmatrix} \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(y_j-\bar{y}) & \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) \\ \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)(y_j-\bar{y}) & \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 \end{vmatrix}}{\begin{vmatrix} \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2 & \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) \\ \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) & \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 \end{vmatrix}} =$$

$$= \frac{\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(y_j-\bar{y})\sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 - \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)(y_j-\bar{y})\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)}{\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2\sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 - \left[\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)\right]^2} \quad (16.9)$$

$$a_2 = \frac{\begin{vmatrix} \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2 & \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(y_j-\bar{y}) \\ \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) & \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)(y_j-\bar{y}) \end{vmatrix}}{\begin{vmatrix} \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2 & \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) \\ \sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2) & \sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2 \end{vmatrix}} = \quad (16.10)$$

$$= \frac{\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2\sum_{j=1}^{n}(x_{2j}-\bar{x}_2)(y_j-\bar{y})-\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(y_j-\bar{y})}{\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)^2\sum_{j=1}^{n}(x_{2j}-\bar{x}_2)^2-\left[\sum_{j=1}^{n}(x_{1j}-\bar{x}_1)(x_{2j}-\bar{x}_2)\right]^2}$$

Sistemul (16.8) se poate rezolva utilizând metodele clasice: metoda simplex, metoda lui *Gauss*, regula lui *Cramer*, schema lui *Doolittle* şi altele.

16.2.2. Intensitatea corelaţiei multiple liniare

Pentru măsurarea intensităţii corelaţiei multiple liniare se poate utiliza raportul de corelaţie. În condiţiile legăturii dintre k variabile independente $X_1, X_2, \ldots, X_k$ şi o variabilă dependentă Y, raportul de corelaţie se determină cu ajutorul formulei:

$$\eta_{y.x_1\cdots x_k} = \sqrt{\frac{\sigma^2_{y.x_1x_2\cdots x_k}}{\sigma^2_y}} = \sqrt{1-\frac{\sum_{j=1}^{n}(y_j - y_{x_1x_2\cdots x_k})^2}{\sum_{j=1}^{n}(y_j-\bar{y})^2}} \qquad (16.11)$$

În cazul a două variabile independente, raportul de corelaţie multiplă va fi:

$$\eta_{y.x_1x_2} = \sqrt{1-\frac{\sum_{j=1}^{n}(y_j - y_{x_1x_2})^2}{\sum_{j=1}^{n}(y_j-\bar{y})^2}} \qquad (16.12)$$

Dacă se cunosc parametrii ecuaţiei de regresie multiplă liniară $a_0, a_1, a_2, \ldots, a_i, \ldots, a_k$ atunci calculul raportului de corelaţie se poate simplifica.

Sumele de pătrate ale abaterilor din formula (16.11) devin:

Scanned with OKEN Scanner

$$\sum_{j=1}^{n}(y_j - y_{x_1x_2\cdots x_k})^2 = \sum_{j=1}^{n}(y_j - y_{x_1x_2\cdots x_k})y_j \qquad (16.13)$$

$$\sum_{j=1}^{n}(y_j - y_{x_1x_2\cdots x_k})^2 = \sum_{j=1}^{n}y_j^2 - a_0\sum_{j=1}^{n}y_j - a_1\sum_{j=1}^{n}x_{1j}y_j - a_2\sum_{j=1}^{n}x_{2j}y_j - \cdots - a_k\sum_{j=1}^{n}x_{kj}y_j$$

iar numitorul

$$\sum_{j=1}^{n}(y_j - \bar{y})^2 = \sum_{j=1}^{n}y_j^2 - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n} \qquad (16.14)$$

Substituind în relaţia (16.12) pe (16.13) şi (16.14), rezultă:

$$\eta_{y.x_1\cdots x_k} = \sqrt{\frac{a_0\sum_{j=1}^{n}y_j + a_1\sum_{j=1}^{n}x_{1j}y_j + \cdots + a_i\sum_{j=1}^{n}x_{ij}y_j + \cdots + a_k\sum_{j=1}^{n}x_{kj}y_j - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}{\sum_{j=1}^{n}y_j^2 - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}} \qquad (16.15)$$

În cazul a două variabile independente X_1 şi X_2 şi o singură variabilă dependentă Y, raportul de corelaţie multiplă liniară devine:

$$\eta_{y.x_1x_2} = \sqrt{\frac{a_0\sum_{j=1}^{n}y_j + a_1\sum_{j=1}^{n}x_{1j}y_j + a_2\sum_{j=1}^{n}x_{2j}y_j - \frac{\left(\sum_{j=1}^{n}y_j^2\right)^2}{n}}{\sum_{j=1}^{n}y_j^2 - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}} \qquad (16.16)$$

Scanned with OKEN Scanner

iar în cazul legăturii dintre Y şi X_1, X_2, X_3 raportul de corelaţie multiplă liniară va fi:

$$\eta_{y.x_1x_2x_3} = \sqrt{\frac{a_0\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} x_{1j}y_j + a_2\sum_{j=1}^{n} x_{2j}y_j + a_3\sum_{j=1}^{n} x_{3j}y_j - \frac{\left(\sum_{j=1}^{n} y_j^2\right)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}}} \quad (16.17)$$

Formula (16.17) se utilizează în practică, atunci când datele empirice, cu privire la variabilele luate în studiu, se prezintă sub forma unor şiruri de valori paralele neponderate.

În cazurile al doilea şi al treilea, formulele raportului de corelaţie multiplă, va trebui să se ţină seama de frecvenţele corespunzătoare ale perechilor de valori corelate.

Intensitatea corelaţiei multiple liniare se poate măsura şi cu ajutorul *coeficientului de corelaţie multiplă*, calculat pe baza coeficienţilor de corelaţie simplă dintre variabilele perechi. Astfel, în cazul corelaţiei dintre o variabilă factorială Y şi două variabile independente X_1, X_2, coeficientul de corelaţie multiplă liniară - notat cu $R_{y.x_1x_2}$, se poate calcula după formula:

$$R_{y.x_1x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}} \quad (16.18)$$

în care:

$$r_{yx_1} = \frac{n\sum_{j=1}^{n} x_1y - \sum_{j=1}^{n} x_1\sum_{j=1}^{n} y}{\sqrt{\left[n\sum_{j=1}^{n} x_1^2 - \left(\sum_{j=1}^{n} x_1\right)^2\right]\left[n\sum_{j=1}^{n} y^2 - \left(\sum_{j=1}^{n} y\right)^2\right]}}$$

Scanned with OKEN Scanner

$$r_{yx_2} = \frac{n\sum_{j=1}^{n} x_2 y - \sum_{j=1}^{n} x_2 \sum_{j=1}^{n} y}{\sqrt{\left[n\sum_{j=1}^{n} x_2^2 - \left(\sum_{j=1}^{n} x_2\right)^2\right]\left[n\sum_{j=1}^{n} y^2 - \left(\sum_{j=1}^{n} y\right)^2\right]}}$$

$$r_{x_1x_2} = \frac{n\sum_{j=1}^{n} x_1 x_2 - \sum_{j=1}^{n} x_1 \sum_{j=1}^{n} x_2}{\sqrt{\left[n\sum_{j=1}^{n} x_1^2 - \left(\sum_{j=1}^{n} x_1\right)^2\right]\left[n\sum_{j=1}^{n} x_2^2 - \left(\sum_{j=1}^{n} x_2\right)^2\right]}}$$

Coeficientul de corelaţie multiplă liniară se poate determina şi cu ajutorul unei alte formule bazată pe determinanţii formaţi din coeficienţii de corelaţie simplă liniară, astfel:

- în cazul a k+1 variabile se vor obţine determinanţii:

$$\Delta yy_{x_1x_2\cdots x_k} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_i} & \cdots & r_{yx_k} \\ r_{x_1y} & 1 & r_{x_1x_2} & \cdots & r_{x_1x_i} & \cdots & r_{x_1x_k} \\ r_{x_2y} & r_{x_2x_1} & 1 & \cdots & r_{x_2x_i} & \cdots & r_{x_2x_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} & \cdots & r_{x_kx_i} & \cdots & 1 \end{vmatrix} \quad (16.19)$$

şi

$$\Delta_y = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_i} & \cdots & r_{x_1x_k} \\ r_{x_2x_k} & 1 & r_{x_2x_3} & \cdots & r_{x_2x_i} & \cdots & r_{x_2x_k} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \cdots & r_{x_kx_i} & \cdots & 1 \end{vmatrix}$$

- în cazul a trei variabile:

Scanned with OKEN Scanner

$$R_{y.x_1x_2} = \sqrt{1-\frac{\Delta yy_{x_1x_2}}{\Delta y}} \tag{16.20}$$

- în cazul a patru variabile:

$$R_{y.x_1x_2x_3} = \sqrt{1-\frac{\Delta yy_{x_1x_2x_3}}{\Delta y}} \tag{16.21}$$

- în cazul general a $k+1$ variabile:

$$R_{y.x_1x_2\cdots x_k} = \sqrt{1-\frac{\Delta yy_{x_1x_2\cdots x_k}}{\Delta y}} \tag{16.22}$$

16.3. Corelaţia parţială

În cazul corelaţiei simple, se studiază legătura dintre doi factori, neglijându-se influenţa celorlalţi factori care acţionează concomitent asupra variabilei rezultative Y, în timp ce, în cazul corelaţiei multiple liniare, se studiază influenţa simultană a două sau mai multe variabile factoriale asupra celei rezultative.

O altă problemă ce se poate rezolva cu ajutorul metodei regresiei şi corelaţiei este determinarea influenţei unei variabile independente asupra variabilei dependente, presupunând că celelalte variabile independente se menţin la un nivel constant. Indicatorul intensităţii unei astfel de corelaţii poartă denumirea de *coeficient de corelaţie parţială*.

Coeficienţii de corelaţie parţială, în cazul a două variabile independente şi o singură variabilă dependentă, se calculează cu ajutorul formulelor:

- între Y şi X_1, neglijând influenţa lui X_2:

$$r_{yx_1.x_2} = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{\sqrt{(1-r_{yx_2}^2)}\sqrt{(1-r_{x_1x_2}^2)}} \tag{16.23}$$

- între Y şi X_2, neglijând influenţa lui X_1:

$$r_{yx_2.x_1} = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{\sqrt{(1-r_{yx_1}^2)}\sqrt{(1-r_{x_1x_2}^2)}} \tag{16.24}$$

Scanned with OKEN Scanner

Coeficienţii de corelaţie parţială (16.23) şi (16.24) poartă denumirea de *coeficienţi de corelaţie parţială de ordinul întâi.*

- între Y şi X_1, neglijând influenţa lui X_2 şi X_3:

$$r_{yx_1 \cdot x_2 x_3} = \frac{r_{yx_1 \cdot x_2} - r_{yx_3 \cdot x_2} r_{x_1 x_3 \cdot x_2}}{\sqrt{(1 - r^2_{yx_3 \cdot x_2})}\sqrt{(1 - r^2_{x_1 x_3 \cdot x_2})}} \tag{16.25}$$

- între Y şi X_2 neglijând influenţa lui X_1 şi X_3:

$$r_{yx_2 \cdot x_1 x_3} = \frac{r_{yx_2 \cdot x_1} - r_{yx_3 \cdot x_1} r_{x_2 x_3 \cdot x_1}}{\sqrt{(1 - r^2_{yx_3 \cdot x_1})}\sqrt{(1 - r^2_{x_2 x_3 \cdot x_1})}} \tag{16.26}$$

- între Y şi X_3 neglijând influenţa lui X_1 şi X_2:

$$r_{yx_3 \cdot x_1 x_2} = \frac{r_{yx_3 \cdot x_1} - r_{yx_2 \cdot x_1} r_{x_2 x_3 \cdot x_1}}{\sqrt{(1 - r^2_{yx_2 \cdot x_1})}\sqrt{(1 - r^2_{x_2 x_3 \cdot x_1})}} \tag{16.27}$$

În cazul general:

$$r_{yx_1 \cdot x_2 x_3 \ldots x_k} = \frac{r_{yx_1 \cdot x_2 x_3 \cdots x_{k-1}} - r_{yx_k \cdot x_2 x_3 \cdots x_{k-1}} \cdot r_{x_1 x_k \cdot x_2 x_3 \cdots x_{k-1}}}{\sqrt{(1 - r^2_{yx_k \cdot x_2 x_3 \cdots x_{k-1}})}\sqrt{(1 - r^2_{x_1 x_k \cdot x_2 x_3 \cdots x_{k-1}})}} \tag{16.28}$$

Calculul indicatorilor de regresie şi corelaţie multiplă şi parţială, chiar în cazul legăturilor liniare este foarte complicat şi necesită utilizarea calculatoarelor electronice. Volumul calculelor şi gradul de complexitate cresc în proporţie geometrică odată cu creşterea numărului variabilelor corelate în progresie aritmetică.

Scanned with OKEN Scanner

16.4. Determinaţie multiplă şi parţială

Utilizarea metodei corelaţiei în statistică oferă şi posibilitatea determinării contribuţiei fiecărei variabile independente asupra variabilei dependente, ca şi influenţa simultană a tuturor variabilelor independente asupra variabilei dependente. Gradul de influenţă, în primul caz, se determină cu *ajutorul coeficienţilor de determinaţie parţială*, iar în al doilea, cu *coeficientul de determinaţie multiplă*.

Coeficientul de determinaţie multiplă se poate calcula prin ridicarea la pătrat a coeficientului sau raportului corelaţiei multiple.

În cazul a trei variabile, coeficientul de determinaţie se poate obţine pe baza coeficientului de corelaţie multiplă, calculat cu ajutorul coeficienţilor de corelaţie simplă:

$$R^2_{y.x_1x_2} = \frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} r_{yx_2} r_{x_1x_2}}{1 - r^2_{x_1x_2}} \qquad (16.29)$$

Din formula (16.29), rezultă că coeficientul determinaţiei multiple este mai mare decât oricare dintre coeficienţii determinaţiei simple:

$$R^2_{y.x_1x_2} > r^2_{yx_1}$$

$$R^2_{y.x_1x_2} > r^2_{yx_2}$$

Coeficientul determinaţiei multiple $R^2_{y.x_1x_2}$ arată care este influenţa simultană a factorilor cauzali (luaţi în calcul) asupra fenomenului efect şi variază în limitele 0 şi 1.

$$0 < R^2_{y.x_1x_2} < 1 \qquad (16.30)$$

Pentru a se interpreta mai uşor, $R^2_{y.x_1x_2}$ se poate exprima în procente.

Dacă Y este independent de X_1 şi X_2, atunci coeficientul determinaţiei multiple este nul: $R^2_{y.x_1x_2} = 0$, când $r_{yx_1} = 0$ şi $r_{yx_2} = 0$.

În cazul legăturilor funcţionale între Y, X_1 şi X_2 coeficientul determinaţiei totale este egal cu 1. Aceasta înseamnă că $r_{x_1x_2} = 0$, 6iar formula (16.29) devine:

$$R^2_{y.x_1x_2} = r^2_{yx_1} + r^2_{yx_2} \tag{16.31}$$

Coeficienţii de determinaţie parţială se pot determina pe baza formulelor următoare:

- între Y şi X_1 neglijând pe X_2:

$$r^2_{yx_1.x_2} = \frac{R^2_{y.x_1x_2} - r^2_{yx_2}}{1 - r^2_{yx_2}} \tag{16.32}$$

- între Y şi X_2 neglijând pe X_1:

$$r^2_{yx_2.x_1} = \frac{R^2_{y.x_1x_2} - r^2_{yx_1}}{1 - r^2_{yx_1}} \tag{16.33}$$

În cazul a trei variabile calculul coeficienţilor de determinaţie parţială se poate face şi pe baza coeficienţilor de corelaţie parţială (prin ridicarea la pătrat a acestora), cu ajutorul formulelor:

$$r^2_{yx_1.x_2} = (r_{yx_1.x_2})^2 = \frac{(r_{yx_1} - r_{yx_2}r_{x_1x_2})^2}{(1 - r^2_{yx_2})(1 - r^2_{x_1x_2})} \tag{16.34}$$

şi

$$r^2_{yx_2.x_1} = (r_{yx_2.x_1})^2 = \frac{(r_{yx_2} - r_{yx_1}r_{x_1x_2})^2}{(1 - r^2_{yx_1})(1 - r^2_{x_1x_2})} \tag{16.35}$$

Desigur, formulele (16.32) şi (16.33) sunt echivalente cu (14.34) şi respectiv (16.35).

Dacă pe lângă coeficienţii de corelaţie parţială, se cunosc şi mediile aritmetice şi abaterile standard ale variabilelor, atunci se pot determina pe această bază şi coeficienţii ecuaţiei regresiei multiple, cu ajutorul formulelor:

Scanned with OKEN Scanner

$$a_1 = r_{yx_1.x_2} \frac{\sigma_{y.x_1x_2}}{\sigma_{x_1.yx_2}} \tag{16.37}$$

$$a_2 = r_{yx_2.x_1} \frac{\sigma_{y.x_1x_2}}{\sigma_{x_2.yx_1}} \tag{16.38}$$

unde:

$$\sigma_{y.x_1x_2} = \sqrt{\sigma_y^2(1-r_{yx_2}^2)(1-r_{x_1yx_2}^2)}$$

$$\sigma_{x_1.yx_2} = \sqrt{\sigma_{x_1}^2(1-r_{x_1x_2}^2)(1-r_{yx_1.x_2}^2)}$$

$$\sigma_{x_2.yx_1} = \sqrt{\sigma_{x_2}^2(1-r_{x_1x_2}^2)(1-r_{yx_2.x_1}^2)}$$

Coeficientul determinaţiei multiple se poate determina şi cu ajutorul raportului de corelaţie multiplă liniară. Astfel, pentru trei variabile în baza formulei (16.29), se obţine:

$$R_{y.x_1x_2}^2 = \eta_{y.x_1x_2}^2 = 1 - \frac{\sum_{j=1}^{n}(y_j - y_{x_1x_2})^2}{\sum_{j=1}^{n}(y_j - \bar{y})^2} = D_{y.x_1x_2}$$

sau

$$R^2_{y.x_1x_2} = \frac{a_0\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} y_j x_{1j} + a_2\sum_{j=1}^{n} y_j x_{1j} - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}} \qquad (16.39)$$

Coeficienţii determinaţiei parţiale se pot obţine şi din formula (16.34).

Substituind pe a_0 cu expresia lui din prima ecuaţie a sistemului (16.6) rezultă:

$$R^2_{y.x_1x_2} = \frac{\bar{y}\sum_{j=1}^{1} y_j - a_1\bar{x}_1\sum_{j=1}^{n} y_j - a_2\bar{x}_2\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} y_j x_{1j} + a_2\sum_{j=1}^{n} y_j x_{2j} - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}} = \qquad (16.40)$$

$$= \frac{a_1\left(\sum_{j=1}^{n} y_j x_{1j} - \frac{\sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_j}{n}\right) + a_2\left(\sum_{j=1}^{n} y_j x_{2j} - \frac{\sum_{j=1}^{n} x_{2j}\sum_{j=1}^{n} y_j}{n}\right)}{\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}}$$

Separând numitorii se obţine:

$$R^2_{y.x_1x_2} = \frac{a_1\left(n\sum_{j=1}^{n} y_j x_{1j} - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} + \frac{a_2\left(n\sum_{j=1}^{n} y_j x_{2j} - \sum_{j=1}^{n} x_{2j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \tag{16.41}$$

În formula (16.41), termenii din membrul al doilea reprezintă cei doi coeficienţi de determinaţie parţială:

- între y şi x_1:

$$r^2_{yx_1.x_2} = \frac{a_1\left(n\sum_{j=1}^{n} y_j x_{1j} - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \tag{16.42}$$

- între y şi x_2:

$$r^2_{yx_2x_1} = \frac{a_2\left(n\sum_{j=1}^{n} y_j x_{2j} - \sum_{j=1}^{n} x_{2j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \tag{16.43}$$

Deci

$$R^2_{y.x_1x_2} = r^2_{yx_1.x_2} + r^2_{yx_2.x_1} \tag{16.44}$$

sau

$$D_{y.x_1x_2} = d_{yx_1.x_2} + d_{yx_2.x_1}$$

Scanned with OKEN Scanner

Din relaţia (16.44) rezultă că suma coeficienţilor determinaţiei parţiale este egală cu coeficientul determinaţiei multiple liniare.

Pe măsură ce în modelul de corelaţie se adaugă noi variabile, coeficientul corelaţiei multiple liniare (ca şi coeficientul determinaţiei multiple) creşte şi se pierde câte un grad de libertate a variaţiei explicate. Altfel spus, se obţine o tot mai bună estimare a fenomenului efect y. Creşterea numărului de variabile în modelul de regresie determină o creştere mai mult decât proporţională a volumului calculelor. Astfel, în cazul unei probleme de regresie şi corelaţie multiplă dintre patru variabile y, x_1, x_2, x_3 volumul calculelor se măreşte cu mai mult de o treime faţă de aceeaşi problemă pentru trei variabile y, x_1, x_2.

În cazul a patru variabile y, x_1, x_2, x_3 coeficientul determinaţiei totale va fi egal cu:

$$D_{y.x_1x_2x_3} = \frac{a_0\sum_{j=1}^{n}y_j - a_1\sum_{j=1}^{n}y_jx_{1j} + a_2\sum_{j=1}^{n}y_jx_{2j} + a_3\sum_{j=1}^{n}y_jx_{3j} - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}{\sum_{j=1}^{n}y_j^2 - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}} \qquad (16.45)$$

Pentru stabilirea formulelor de calcul ale celor trei coeficienţi de determinaţie parţială, vom substitui în (16.45) pe a_0 cu valoarea lui din prima ecuaţie a sistemului de ecuaţii normale:

$$D_{y.x_1x_2x_3} = \frac{a_1\left(\sum_{j=1}^{n}y_jx_{1j} - \frac{\sum_{j=1}^{n}x_{1j}\sum_{j=1}^{n}y_j}{n}\right) + a_2\left(\sum_{j=1}^{n}y_jx_{2j} - \frac{\sum_{j=1}^{n}x_{2j}\sum_{j=1}^{n}y_j}{n}\right) + a_3\left(\sum_{j=1}^{n}y_jx_{3j} - \frac{\sum_{j=1}^{n}x_{3j}\sum_{j=1}^{n}y_j}{n}\right)}{\sum_{j=1}^{n}y_j^2 - \frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}$$

Scanned with OKEN Scanner

Rezultă:

$$d_{yx_1 \cdot x_2 x_3} = \frac{a_1\left(n\sum_{j=1}^{n} y_j x_{1j} - \sum_{j=1}^{n} x_{1j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \qquad (16.46)$$

$$d_{yx_2 \cdot x_1 x_3} = \frac{a_2\left(n\sum_{j=1}^{n} y_j x_{2j} - \sum_{j=1}^{n} x_{2j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \qquad (16.47)$$

$$d_{yx_3 \cdot x_1 x_2} = \frac{a_3\left(n\sum_{j=1}^{n} y_j x_{3j} - \sum_{j=1}^{n} x_{3j}\sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2} \qquad (16.48)$$

În general, pentru k variabile, coeficienţii de determinaţie se vor calcula astfel:

$$D_{y.x_1 x_2 \ldots x_k} = \frac{a_0\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} y_j x_{1j} + a_2\sum_{j=1}^{n} y_j x_{2j} + \cdots + a_k\sum_{j=1}^{n} y_j x_{kj} - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{\left(\sum_{j=1}^{n} y_j\right)^2}{n}} \qquad (16.49)$$

Scanned with OKEN Scanner

$$d_{yx_i \cdot x_1 x_2 \cdots x_k} = \frac{a_i \left(n \sum_{j=1}^{n} y_j x_{ij} - \sum_{j=1}^{n} x_{ij} \sum_{j=1}^{n} y_j \right)}{n \sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j \right)^2} \qquad i = \overline{1,k} \qquad (16.50)$$

Între coeficientul determinaţiei totale (16.49) şi coeficienţii determinaţiei parţiale (14.50) există următoarea relaţie: suma coeficienţilor de determinaţie parţială este egală cu coeficientul determinaţiei totale:

$$D_{y \cdot x_1 x_2 \ldots x_k} = \sum_{i=1}^{k} d_{yx_1 \cdot x_2 \ldots x_k} \qquad (16.51)$$

Scanned with OKEN Scanner

17. REGRESIE ŞI CORELAŢIE MULTIPLĂ CURBILINIE

17.1. Regresie multiplă curbilinie

Am văzut anterior că regresia multiplă liniară se caracterizează prin faptul că fenomenul efect variază în mod uniform atunci când fenomenele cauzale se modifică cu o unitate, indiferent de mărimea fiecărei variabile independente. În realitate, dezvoltarea fenomenelor reale urmează rareori modelul liniar. În evoluţia lor se poate identifica de multe ori un nivel optim de creştere, care corespunde unor anumite valori de extrem ale variabilelor cauzale. Modelul liniar nu poate descrie o astfel de evoluţie şi de aceea este necesar să se recurgă la modele curbilinii.

În forma cea mai simplă, o problemă de regresie şi corelaţie multiplă curbilinie se formulează astfel: fie o variabilă dependentă Y care este influenţată de k variabile cauzale, independente $X_1, X_2, \ldots, X_k$, iar legătura dintre variabile este de forma:

$$Y = a + f_1(X_1) + \cdots + f_i(X_i) + \cdots + f_k(X_k) + \varepsilon \qquad (17.1)$$

Relaţia (17.1) arată că variaţia fenomenului efect Y este rezultatul unei sume de funcţii $f(X_i)$ $i = \overline{1,k}$ şi de alţi factori, în afară de X_i, ale căror influenţă totală este egală cu a . Desigur, o parte din funcţiile $f(X_i)$ pot să fie liniare, iar o altă parte, curbilinii. Pentru ca legătura multiplă de corelaţie să fie curbilinie, este suficient ca cel puţin una dintre funcţiile $f(X_i)$ să fie de formă curbilinie.

Determinarea parametrilor ecuaţiei regresiei multiple curbilinii se poate face cu ajutorul *metodei celor mai mici pătrate* sau pe baza *metodei aproximării grafice succesive*.

Modelul regresiei multiple curbilinii este necesar să se exprime printr-o funcţie, cât mai simplă din punct de vedere matematic, în aşa fel încât parametrii ecuaţiei de regresie să se poată determina cu ajutorul unor metode obişnuite şi să fie

Scanned with OKEN Scanner

uşor de interpretat. Dacă, pentru legăturile simple dintre Y şi X_i se utilizează un model de forma polinomului de gradul doi, atunci ecuaţia regresiei multiple curbilinii (17.1) va deveni:

$$Y = a + a_1 X_1 + b_1 X_1^2 + a_2 X_2 + b_2 X_2^2 + \cdots + a_k X_k + b_k X_k^2 \qquad (17.2)$$

în care: $f(X_i) = a_0 + a_i X_i + b_i X_i^2 \qquad i = \overline{1,k}$

În practică se pot distinge două situaţii diferite:

1) când se cunosc sau se pot stabili, pe baza analizei statistice, formele ecuaţiilor simple de regresie dintre variabila efect şi fiecare variabilă cauzală, se poate determina o funcţie de regresie multiplă curbilinie corespunzătoare, ai cărei parametri se pot estima cu ajutorul metodei celor mai mici pătrate;

2) dacă nu se cunoaşte forma reală a legăturii dintre variabila dependentă şi fiecare variabilă independentă, atunci se poate proceda la o reprezentare grafică cât mai adecvată a legăturilor simple şi, analizând esenţa fenomenelor corelate prin încercări succesive, se ajunge la stabilirea celei mai corespunzătoare forme a ecuaţiei de regresie simplă curbilinie şi apoi se va elabora modelul regresiei multiple curbilinii.

$$\begin{cases}
na_0 + a_1\sum_{j=1}^{n} x_1 + b_1\sum_{j=1}^{n} x_1^2 + a_2\sum_{j=1}^{n} x_2 + b_2\sum_{j=1}^{n} x_2^2 + \cdots + a_k\sum_{j=1}^{n} x_k + b_k\sum_{j=1}^{n} x_k^2 = \sum_{j=1}^{n} y_j \\
a_0\sum_{j=1}^{n} x_1 + a_1\sum_{j=1}^{n} x_1^2 + b_1\sum_{j=1}^{n} x_1^3 + a_2\sum_{j=1}^{n} x_1x_2 + b_2\sum_{j=1}^{n} x_1x_2^2 + \cdots + a_k\sum_{j=1}^{n} x_1x_k + b_k\sum_{j=1}^{n} x_1x_k^2 = \sum_{j=1}^{n} x_1y_j \\
a_0\sum_{j=1}^{n} x_1^2 + a_1\sum_{j=1}^{n} x_1^3 + b_1\sum_{j=1}^{n} x_1^4 + a_2\sum_{j=1}^{n} x_1^2x_2 + b_2\sum_{j=1}^{n} x_1^2x_2^2 + \cdots + a_k\sum_{j=1}^{n} x_1^2x_k + b_k\sum_{j=1}^{n} x_1^2x_k^2 = \sum_{j=1}^{n} x_1^2y_j \\
a_0\sum_{j=1}^{n} x_2 + a_1\sum_{j=1}^{n} x_1x_2 + b_1\sum_{j=1}^{n} x_1^2x_2 + a_2\sum_{j=1}^{n} x_2^2 + b_2\sum_{j=1}^{n} x_2^3 + \cdots + a_k\sum_{j=1}^{n} x_2x_k + b_k\sum_{j=1}^{n} x_2x_k^2 = \sum_{j=1}^{n} x_2y_j \\
\cdots\cdots\cdots

Parametrii ecuaţiei regresiei multiple (17.2) de tip polinomial se pot determina prin aplicarea metodei celor mai mici pătrate, cu ajutorul sistemului de ecuaţii normale (17.2).

În sistemul (17.3) şi în tot capitolul 17, fiecare variabilă X_i variază după indicele j. Pentru a simplifica simbolizarea, nu s-a mai scris cel de-al doilea indice.

În sistemul 17.3 fiecare variabilă variază după j şi însumările se fac după acest indice $j = \overline{1,n}$, care pentru X_i se subînţelege.

Pentru a prezenta mai detaliat metodele de regresie şi corelaţie multiplă curbilinie să considerăm cazul mai simplu al legăturii de tip parabolic între două variabile independente X_1 şi X_2 şi o singură variabilă dependentă Y.

Conform relaţiei (17.2), modelul regresiei multiple curbilinii pentru trei variabile corelate va fi descris de următoarea ecuaţie:

$$Y = a + a_1X_1 + b_1X_1^2 + a_2X_2 + b_2X_2^2 \qquad (17.4)$$

Parametrii acestei ecuaţii de regresie se pot determina tot cu ajutorul metodei celor mai mici pătrate, pe baza sistemului:

$$\begin{cases} na + a_1\sum_{j=1}^{n}x_{1j} + b_1\sum_{j=1}^{n}x_{1j}^2 + a_2\sum_{j=1}^{n}x_{2j} + b_2\sum_{j=1}^{n}x_{2j}^2 = \sum_{j=1}^{n}y_j \\ a\sum_{j=1}^{n}x_{1j} + a_1\sum_{j=1}^{n}x_{1j}^2 + b_1\sum_{j=1}^{n}x_{1j}^3 + a_2\sum_{j=1}^{n}x_{1j}x_{2j} + b_2\sum_{j=1}^{n}x_{1j}x_{2j}^2 = \sum_{j=1}^{n}y_jx_{1j} \\ a\sum_{j=1}^{n}x_{1j}^2 + a_1\sum_{j=1}^{n}x_{1j}^3 + b_1\sum_{j=1}^{n}x_{1j}^4 + a_2\sum_{j=1}^{n}x_{1j}^2x_{2j} + b_2\sum_{j=1}^{n}x_{1j}^2x_{2j}^2 = \sum_{j=1}^{n}y_jx_{1j}^2 \\ a\sum_{j=1}^{n}x_{2j} + a_1\sum_{j=1}^{n}x_{1j}x_{2j} + b_1\sum_{j=1}^{n}x_{1j}^2x_{2j} + a_2\sum_{j=1}^{n}x_{2j}^2 + b_2\sum_{j=1}^{n}x_{2j}^3 = \sum_{j=1}^{n}y_jx_{2j} \\ a\sum_{j=1}^{n}x_{2j}^2 + a_1\sum_{j=1}^{n}x_{1j}x_{2j}^2 + b_1\sum_{j=1}^{n}x_{1j}^2x_{2j}^2 + a_2\sum_{j=1}^{n}x_{2j}^3 + b_2\sum_{j=1}^{n}x_{2j}^4 = \sum_{j=1}^{n}y_jx_{2j}^2 \end{cases} \qquad (17.5)$$

Scanned with OKEN Scanner

Prin rezolvarea sistemului (17.5) se vor determina parametrii ecuaţiei regresiei multiple curbilinii: a_1, a_2, b_1, b_2.

Reprezentarea grafică a ecuaţiei regresiei multiple curbilinii se face printr-o suprafaţă în spaţiu, ca şi în cazul legăturilor multiple liniare, dar se deosebeşte de acestea prin faptul că ea nu mai este o suprafaţă plană (fig.17.1.).

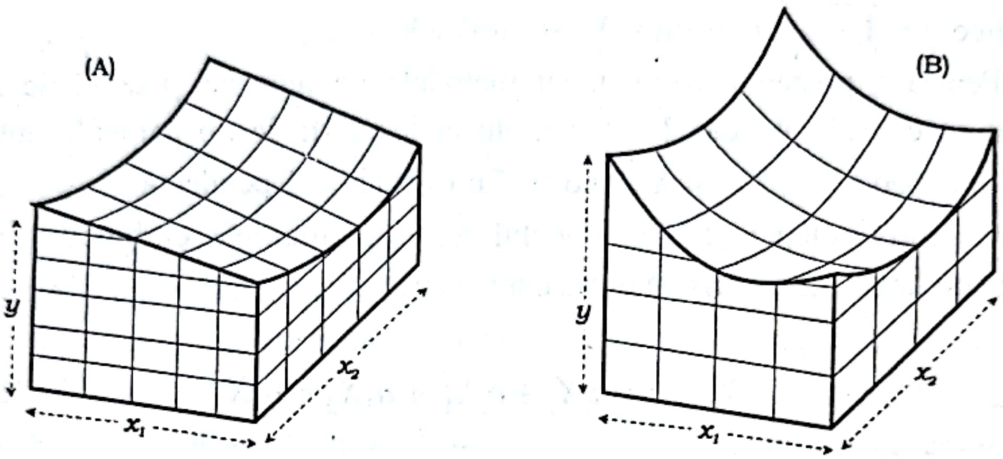


Figura 17.1. Legături multiple curbilinii între trei variabile

În figura 17.1. (A), se prezintă o legătură multiplă dintre y, x_1 şi x_2 în cazul în care legăturile simple dintre variabile au forme diferite, astfel între y şi x_1 există o legătură curbilinie, iar între y şi x_2 o legătură liniară, iar în figura 17.1 (B) se prezintă sub forma suprafeţei de regresie, în cazul în care regresia simplă dintre y şi x_1, cât şi cea dintre y şi x_2 sunt curbilinii.

17.1.1. Liniarizarea legăturilor multiple curbilinii

Creşterea economică, în anumite condiţii, se poate estima, după cum se ştie, cu ajutorul modelului descris de o funcţie de tip *Cobb-Douglas*:

$$Y_X = aX_1^{b_1} X_2^{b_2}$$

Această funcţie se poate *liniariza prin logaritmare* şi devine:

Scanned with OKEN Scanner

$$lg\ y_x = lg\ a + b_1\ lg\ x_1 + b_2\ lg\ x_2 \qquad (17.6)$$

Forma liniarizată se pretează cu uşurinţă la aplicarea metodei celor mai mici pătrate, astfel:

$$S = \sum_{j=1}^{n} (lg\ y_j - lg\ y_{x_i})^2 = minim$$

sau

$$S = \sum_{j=1}^{n} (lg\ y_j - lg\ a - b_1\ lg\ x_{1j} - b_2\ lg\ x_{2j})^2 = minim$$

Calculând derivatele parţiale şi egalându-le cu zero, rezultă următorul sistem de ecuaţii normale:

$$\begin{cases} n\lg a + b_1 \sum_{j=1}^{n} \lg x_{1j} + b_2 \sum_{j=1}^{n} \lg x_{2j} = \sum_{j=1}^{n} \lg y_j \\ \lg a \sum_{j=1}^{n} \lg x_{1j} + b_1 \sum_{j=1}^{n} (\lg x_{1j})^2 + b_2 \sum_{j=1}^{n} \lg x_{2j} \lg x_{1j} = \sum_{j=1}^{n} \lg y_j \lg x_{1j} \\ \lg a \sum_{j=1}^{n} \lg x_{2j} + b_1 \sum_{j=1}^{n} \lg x_{1j} \lg x_{2j} + b_2 \sum_{j=1}^{n} (\lg x_{2j})^2 = \sum_{j=1}^{n} \lg y_j \lg x_{2j} \end{cases} \qquad (17.7)$$

Sistemul (17.7) se poate rezolva prin metodele clasice de rezolvare a sistemelor de ecuaţii normale.

În cazul regresiei multiple curbilinii descrise de modelul (17.3), liniarizarea se obţine tot prin logaritmare, astfel:

$$lg\ y_x = lg\ a + b_1\ lg\ x_1 + b_2\ lg\ x_2 + b_3\ lg\ x_3 \qquad (17.8)$$

Parametrii modelului (17.8) se pot determina de asemenea, cu ajutorul metodei celor mai mici pătrate, pe baza sistemului:

Scanned with OKEN Scanner

$$\begin{cases} n\lg a + b_1\sum_{j=1}^{n}\lg x_{1j} + b_2\sum_{j=1}^{n}\lg x_{2j} + b_3\sum_{j=1}^{n}\lg x_{3j} = \sum_{j=1}^{n}\lg y_j \\ \lg a\sum_{j=1}^{n}\lg x_{1j} + b_1\sum_{j=1}^{n}(\lg x_{1j})^2 + b_2\sum_{j=1}^{n}\lg x_{2j}\lg x_{1j} + b_3\sum_{j=1}^{n}\lg x_{1j}\lg x_{3j} = \sum_{j=1}^{n}\lg y_j \lg x_{1j} \\ \lg a\sum_{j=1}^{n}\lg x_{2j} + b_1\sum_{j=1}^{n}\lg x_{1j}\lg x_{2j} + b_2\sum_{j=1}^{n}(\lg x_{2j})^2 + b_3\sum_{j=1}^{n}\lg x_{2j}\lg x_{3j} = \sum_{j=1}^{n}\lg y_j \lg x_{2j} \\ \lg a\sum_{j=1}^{n}\lg x_{3j} + b_1\sum_{j=1}^{n}\lg x_{1j}\lg x_{3j} + b_2\sum_{j=1}^{n}\lg x_{2j}\lg x_{3j} + b_3\sum_{j=1}^{n}(\lg x_{3j})^2 = \sum_{j=1}^{n}\lg y_j \lg x_{3j} \end{cases} \quad (17.9)$$

17.2 Intensitatea corelaţiei multiple curbilinii

Ca şi în cazul corelaţiei multiple de formă liniară, intensitatea corelaţiei multiple curbilinii se măsoară cu ajutorul raportului de corelaţie:

$$\eta_{y\cdot x_1x_2\ldots x_k} = \sqrt{\frac{a\sum_{j=1}^{n}y_j + a_1\sum_{j=1}^{n}x_1y_j + b_1\sum_{j=1}^{n}x_1^2y_j + a_2\sum_{j=1}^{n}y_j + b_2\sum_{j=1}^{n}x_2^2y_j + \cdots + a_k\sum_{j=1}^{n}x_ky_j + b_k\sum_{j=1}^{n}x_k^2y_j - \frac{(\sum_{j=1}^{n}y_j)^2}{n}}{\sum_{j=1}^{n}y_j^2 - \frac{(\sum_{j=1}^{n}y_j)^2}{n}}} \quad (17.10)$$

În particular, pentru regresiile curbilinii polinomiale dintre trei variabile raportul de corelaţie (17.10) devine:

Scanned with OKEN Scanner

$$\eta_{y.x_1x_2} = \sqrt{\frac{a\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} x_1y_j + b_1\sum_{j=1}^{n} x_1^2y_j + a_2\sum_{j=1}^{n} x_2y_j + b_2\sum_{j=1}^{n} x_2^2y_j - \frac{(\sum_{j=1}^{n} y_j)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{(\sum_{j=1}^{n} y_j)^2}{n}}} \quad (17.11)$$

În cazul a patru variabile, raportul de corelaţie (17.10) va fi:

$$\eta_{y.x_1x_2x_3} = \sqrt{\frac{a\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} x_1y_j + b_1\sum_{j=1}^{n} x_1^2y_j + b_2\sum_{j=1}^{n} x_2^2y_j + a_3\sum_{j=1}^{n} x_3y_j + b_3\sum_{j=1}^{n} x_3^2y_j - \frac{(\sum_{j=1}^{n} y_j)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{(\sum_{j=1}^{n} y_j)^2}{n}}} \quad (17.12)$$

Intensitatea legăturii dintre variabile, în cazul unui model de corelaţie multiplă de tipul unei funcţii de producţie cu doi factori cauzali, se va determina cu ajutorul raportului de corelaţie multiplă:

$$\eta_{y.x_1x_2} = \sqrt{\frac{lg\, a\sum_{j=1}^{n} y_j + b_1\sum_{j=1}^{n} lg\, x_{1j}\, lg\, y_j + b_2\sum_{j=1}^{n} lg\, x_{2j}\, lg\, y_j - \frac{(\sum_{j=1}^{n} lg\, y_j)^2}{n}}{\sum_{j=1}^{n} (lg\, y_j)^2 - \frac{(\sum_{j=1}^{n} lg\, y_j)^2}{n}}} \quad (17.13)$$

iar în cazul a trei factori cauzali:

Scanned with OKEN Scanner

$$\eta_{y\cdot x_1x_2} = \sqrt{\frac{\lg a\sum_{j=1}^{n} y_j + b_1\sum_{j=1}^{n} \lg x_{1j}\lg y_j + b_2\sum_{j=1}^{n} \lg x_{2j}\lg y_j + b_3\sum_{j=1}^{n} \lg x_{3j}\lg y_j - \frac{(\sum_{j=1}^{n} \lg y_j)^2}{n}}{\sum_{j=1}^{n} (\lg y_j)^2 - \frac{(\sum_{j=1}^{n} \lg y_j)^2}{n}}}$$

(17.14)

17.2.1. Determinaţie curbilinie multiplă şi parţială

Ca şi în cazurile anterioare, coeficientul determinaţiei multiple curbilinii se obţine prin ridicarea la pătrat a raportului de corelaţie:

$$D_{y\cdot x_1x_2\cdots x_k} = \eta^2_{y\cdot x_1x_2\cdots x_k}$$

În particular, în cazul legăturilor curbilinii descrise de modelul polinomului de gradul doi, coeficientul determinaţiei multiple devine:

$$D_{y\cdot x_1x_2\cdots x_k} = \frac{a\sum_{j=1}^{n} y_j + a_1\sum_{j=1}^{n} x_1y_j + b_1\sum_{j=1}^{n} x_1^2y_j + a_2\sum_{j=1}^{n} x_2y_j + b_2\sum_{j=1}^{n} x_2^2y_j + \cdots + a_k\sum_{j=1}^{n} x_ky_j + b_k\sum_{j=1}^{n} x_k^2y_j - \frac{(\sum_{j=1}^{n} y_i)^2}{n}}{\sum_{j=1}^{n} y_j^2 - \frac{(\sum_{j=1}^{n} y_j)^2}{n}}$$

(17.15)

Mărimea influenţei fiecăruia dintre cei k factori, care acţionează asupra fenomenului efect, se poate măsura cu ajutorul coeficienţilor de determinaţie parţială. Formulele lor de calcul se obţin prin substituirea parametrului a în (17.15) cu valoarea lui care rezultă din prima ecuaţie a sistemului de ecuaţii normale (17.3).

Scanned with OKEN Scanner

Astfel :

$$a=\bar{y}-a_1\bar{x}_1-b_1\bar{x}_1^2-a_2\bar{x}_2-b_2\bar{x}_2^2-\cdots-a_i\bar{x}_i-b_i\bar{x}_i^2-\cdots-a_k\bar{x}_k-b_k\bar{x}_k^2$$

$$D_{y\,x_1x_2\ldots x_k}=$$

$$=\frac{\frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}-a_1\frac{\sum_{j=1}^{n}x_1\sum_{j=1}^{n}y_j}{n}-b_1\frac{\sum_{j=1}^{n}x_1^2\sum_{j=1}^{n}y_j}{n}-a_2\frac{\sum_{j=1}^{n}x_2\sum_{j=1}^{n}y_j}{n}-b_2\frac{\sum_{j=1}^{n}x_2^2\sum_{j=1}^{n}y_j}{n}-\cdots-a_k\frac{\sum_{j=1}^{n}x_k\sum_{j=1}^{n}y_j}{n}-b_k\frac{\sum_{j=1}^{n}x_k^2\sum_{j=1}^{n}y_j}{n}+}{\sum_{j=1}^{n}y_j^2-\frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}$$

$$=\frac{+a_1\sum_{j=1}^{n}x_1y_j+b_1\sum_{j=1}^{n}x_1^2y_j+a_2\sum_{j=1}^{n}x_2y_j+b_2\sum_{j=1}^{n}x_2^2y_j+\cdots+a_k\sum_{j=1}^{n}x_ky_j+b_k\sum_{j=1}^{n}x_k^2y_j-\frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}{\sum_{j=1}^{n}y_j^2-\frac{\left(\sum_{j=1}^{n}y_j\right)^2}{n}}=$$

$$=\frac{a_1\left(n\sum_{j=1}^{n}x_1y_j-\sum_{j=1}^{n}x_1\sum_{j=1}^{n}y_j\right)+b_1\left(n\sum_{j=1}^{n}x_1^2y_j-\sum_{j=1}^{n}x_1^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}+$$

$$+\frac{a_2\left(n\sum_{j=1}^{n}x_2y_j-\sum_{j=1}^{n}x_2\sum_{j=1}^{n}y_j\right)+b_2\left(n\sum_{j=1}^{n}x_2^2y_j-\sum_{j=1}^{n}x_2^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}+\cdots$$

Scanned with OKEN Scanner

$$+\cdots+\frac{a_i\left(n\sum_{j=1}^{n}x_iy_j-\sum_{j=1}^{n}x_i\sum_{j=1}^{n}y_j\right)+b_i\left(n\sum_{j=1}^{n}x_i^2y_j-\sum_{j=1}^{n}x_i^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}+\cdots$$

$$+\cdots+\frac{a_k\left(n\sum_{j=1}^{n}x_ky_j-\sum_{j=1}^{n}x_k\sum_{j=1}^{n}y_j\right)+b_k\left(n\sum_{j=1}^{n}x_{k2}^2y_j-\sum_{j=1}^{n}x_k^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}=$$

$$=d_{y.x_1x_2x_3\ldots x_k}+d_{yx_2.x_1x_3\ldots x_k}+\cdots+d_{yx_i.x_1x_2\ldots x_k}+\cdots+d_{yx_k.x_1x_2\ldots x_{k-1}} \qquad (17.16)$$

Din relaţia (17.16), rezultă formulele coeficienţilor de determinaţie parţială, în cazul legăturii curbilinii luate în calcul:

$$d_{yx_1.x_2x_3\ldots x_k}=\frac{a_1\left(n\sum_{j=1}^{n}x_1y_j-\sum_{j=1}^{n}x_1\sum_{j=1}^{n}y_j\right)+b_1\left(n\sum_{j=1}^{n}x_1^2y_j-\sum_{j=1}^{n}x_1^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2}$$

$$d_{yx_2.x_1x_3\ldots x_k}=\frac{a_2\left(n\sum_{j=1}^{n}x_2y_j-\sum_{j=1}^{n}x_2\sum_{j=1}^{n}y_j\right)+b_2\left(n\sum_{j=1}^{n}x_2^2y_j-\sum_{j=1}^{n}x_2^2\sum_{j=1}^{n}y_j\right)}{n\sum_{j=1}^{n}y_j^2-\left(\sum_{j=1}^{n}y_j\right)^2} \qquad (17.17)$$

.....

Scanned with OKEN Scanner

$$d_{yx_i.x_1x_2...x_k} = \frac{a_i\left(n\sum_{j=1}^{n} x_i y_j - \sum_{j=1}^{n} x_i \sum_{j=1}^{n} y_j\right) + b_i\left(n\sum_{j=1}^{n} x_i^2 y_j - \sum_{j=1}^{n} x_i^2 \sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}$$

.....

$$d_{yx_k.x_1x_2...x_{k-1}} = \frac{a_k\left(n\sum_{j=1}^{n} x_k y_j - \sum_{j=1}^{n} x_k \sum_{j=1}^{n} y_j\right) + b_k\left(n\sum_{j=1}^{n} x_k^2 y_j - \sum_{j=1}^{n} x_k^2 \sum_{j=1}^{n} y_j\right)}{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}$$

Deci,

$$D_{y.x_1x_2...x_k} = d_{yx_1.x_2x_3...x_k} + d_{yx_2.x_1x_3...x_k} + \cdots + d_{yx_i.x_1x_2...x_k} + \cdots + d_{yx_k.x_1x_2...x_{k-1}} \qquad (17.18)$$

În cazul unei funcţii de producţie cu doi factori cauzali, coeficienţii de determinaţie vor lua următoarele expresii:

$$D_{y.x_1x_2} = \frac{\lg a\sum_{j=1}^{n} \lg y_j + b_1\sum_{j=1}^{n} \lg x_{1j} \lg y_j + b_2\sum_{j=1}^{n} \lg x_{2j} \lg y_j - \frac{\left(\sum_{j=1}^{n} \lg y_j\right)^2}{n}}{\sum_{j=1}^{n} (\lg y_j)^2 - \frac{\left(\sum_{j=1}^{n} \lg y_j\right)^2}{n}} \qquad (17.19)$$

şi

Scanned with OKEN Scanner

$$d_{yx_1 \cdot x_2} = \frac{b_1\left(n\sum_{j=1}^{n} lg\, y_j\, lg\, x_{1j} - \sum_{j=1}^{n} lg\, y_j \sum_{j=1}^{n} lg\, x_{1j}\right)}{n\sum_{j=1}^{n}(lg\, y_j)^2 - \left(\sum_{j=1}^{n} lg\, y_j\right)^2} \qquad (17.20)$$

$$d_{yx_2 \cdot x_1} = \frac{b_2\left(n\sum_{j=1}^{n} lg\, y_j\, lg\, x_{2j} - \sum_{j=1}^{n} lg\, y_j \sum_{j=1}^{n} lg\, x_{2j}\right)}{n\sum_{j=1}^{n}(lg\, y_j)^2 - \left(\sum_{j=1}^{n} lg\, y_j\right)^2} \qquad (17.21)$$

iar

$$D_{y \cdot x_1 x_2} = d_{yx_1 \cdot x_2} + d_{yx_2 \cdot x_1} \qquad (17.22)$$

Coeficienţii de determinaţie pentru o funcţie de producţie cu trei factori cauzali se vor determina în mod analog:

$$D_{y \cdot x_1 x_2 x_3} = \frac{lga\sum_{j=1}^{n} lgy_j + b_1\sum_{j=1}^{n} lgx_{1j}\, lgy_j + b_2\sum_{j=1}^{n} lgx_{2j}\, lgy_j + b_3\sum_{j=1}^{n} lgx_{3j}\, lgy_j - \frac{\left(\sum_{j=1}^{n} lgy_j\right)^2}{n}}{\sum_{j=1}^{n}(lgy_j)^2 - \frac{\left(\sum_{j=1}^{n} lgy_j\right)^2}{n}} \qquad (17.23)$$

şi

$$d_{yx_1 \cdot x_2 x_3} = \frac{b_1\left(n\sum_{j=1}^{n} lg\, y_j\, lg\, x_{1j} - \sum_{j=1}^{n} lg\, y_j \sum_{j=1}^{n} lg\, x_{1j}\right)}{n\sum_{j=1}^{n}(lg\, y_j)^2 - \left(\sum_{j=1}^{n} lg\, y_j\right)^2} \qquad (17.24)$$

Scanned with OKEN Scanner

$$d_{yx_2.x_1x_3} = \frac{b_2\left(n\sum_{j=1}^{n} lg\, y_j\, lg\, x_{2j} - \sum_{j=1}^{n} lg\, y_j \sum_{j=1}^{n} lg\, x_{2j}\right)}{n\sum_{j=1}^{n} (lg\, y_j)^2 - \left(\sum_{j=1}^{n} lg\, y_j\right)^2} \qquad (17.25)$$

$$d_{yx_3.x_1x_2} = \frac{b_3\left(n\sum_{j=1}^{n} lg\, y_j\, lg\, x_{3j} - \sum_{j=1}^{n} lg\, y_j \sum_{j=1}^{n} lg\, x_{3j}\right)}{n\sum_{j=1}^{n} (lg\, y_j)^2 - \left(\sum_{j=1}^{n} lg\, y_j\right)^2} \qquad (17.26)$$

iar

$$D_{y.x_1x_2x_3} = d_{yx_1.x_2x_3} + d_{yx_2.x_1x_3} + d_{yx_3.x_1x_2} \qquad (17.27)$$

Scanned with OKEN Scanner

18. CORELAŢIA NEPARAMETRICĂ

18.1. Noţiuni de corelaţie neparametrică

Metodele de studiere a corelaţiei dintre fenomene, prezentate până acum, se aplică în practică numai în cazul variabilelor numerice, normal distribuite sau apropiate de normalitate. Aceste metode poartă denumirea de *metode parametrice.* Dacă distribuţia variabilelor corelate nu este normală sau nu este cunoscută, pentru măsurarea intensităţii corelaţiei dintre ele, se pot utiliza *metode neparametrice.* Metodele neparametrice au fost introduse în statistică de către *C. Spearman* şi au fost dezvoltate de către *K.Pearson, G.Fechner, M.Kendall, N.Georgescu Roegen* şi alţii.

În utilizarea metodelor neparametrice se face abstracţie de tipul distribuţiilor variabilelor corelate, fapt care le măreşte substanţial cadrul de aplicabilitate şi asupra caracteristicilor nenumerice. Metodele neparametrice nu operează cu valorile reale ale caracteristicilor, ci cu rangurile lor sau cu alte mărimi derivate.

În acest scop unităţile statistice se ordonează crescător sau descrescător, după variantele variabilei factoriale şi alăturat se înscriu nivelurile corespunzătoare ale variabilei dependente, apoi se atribuie câte un rang fiecărei variante.

Dacă între variabilele corelate există o legătură directă, atunci unităţile statistice care s-au clasat pe primele locuri în raport cu caracteristica factorială se vor clasa tot pe primele locuri şi în raport cu caracteristica rezultativă, în timp ce, în cazul unei legături inverse, rangurilor mici ale variabilei factoriale le vor corespunde rangurile mari ale variabilei rezultative.

În primul caz există o concordanţă a rangurilor celor două variabile, iar în cazul al doilea o discordanţă de ranguri. Concordanţa dintre ranguri are loc în cazul unor legături directe, iar discordanţa, în cazul legăturilor inverse.

Astfel, între gradul de industrializare a judeţelor *României* şi gradul de urbanizare există o legătură statistică directă, în sensul că, pe măsură ce creşte gradul de dezvoltare industrială a judeţelor, va creşte şi necesarul de forţă de muncă pentru industrie, fapt care va determina o creştere a gradului de urbanizare. Aceasta înseamnă că o comparaţie a rangurilor judeţelor după cele două caracteristici

Scanned with OKEN Scanner

va da posibilitatea să se măsoare intensitatea legăturii statistice dintre cele două fenomene.

Dintre metodele neparametrice de studiere a corelaţiei dintre fenomene vom prezenta în continuare, metoda lui *Spearman*, metoda lui *Kendall* şi metoda lui *Fechner*.

18.2. Coeficientul de corelaţie a rangurilor al lui *Spearman*

Pornind de la ipoteza logică că, în cazul unei legături funcţionale directe dintre variabile, ar trebui să existe o concordanţă deplină între rangurile celor două variabile corelate şi o discordanţă totală în cazul legăturilor funcţionale inverse, *C.Spearman* a stabilit un coeficient de corelaţie a rangurilor dezvoltat şi caracterizat sub multiple aspecte de către *K.Pearson* şi cunoscut sub următoarea formulă:

$$\theta = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} \qquad -1 \le \theta \le 1 \qquad (18.1)$$

în care:

d_i - reprezintă diferenţele dintre rangurile celor două variabile corelate.

n - numărul perechilor de valori corelate.

Coeficientul lui *Spearman* θ , ca şi coeficientul r_{xy} al lui *Pearson* variază între limitele -1 şi $+1$ cu aceeaşi semnificaţie. În cazul concordanţei depline a rangurilor, toate diferenţele $d_i = x_i - y_i$ sunt nule şi deci, $\theta = +1$; iar în cazul a două serii de ranguri total opuse $\theta = -1$.

Demonstraţia, în cazul rangurilor total discordante, este diferită, după cum n este par sau impar:

Dacă n este impar, adică $n = 2m + 1$, valorile lui d_i vor fi:

$$2m, \quad 2m-2, \quad \ldots, \quad 2, \quad 0, \quad -2, \quad \ldots, \quad -(2m-2), \quad -2m$$

Scanned with OKEN Scanner

iar $\sum_{i=1}^{n} d_i^2 = 2\left[(2m)^2 + (2m-2)^2 + \cdots + 2^2\right] = \frac{8m(m+1)(2m+1)}{6}$

Substituind în (18.1) rezultă:

$$\theta = 1 - \frac{8m(m+1)(2m+1)}{(2m+1)\left[(2m+1)^2 - 1\right]} = 1 - \frac{8m(m+1)}{4m(m+1)} = -1$$

Dacă n este par, deci $n = 2m$, vom avea:

$$\sum_{i=1}^{n} d_i^2 = 2\left[(2m-1)^2 + \cdots + 1^2\right] = \frac{2m}{3}(4m^2 - 1) \qquad \text{iar} \qquad \theta = -1$$

Formula coeficientului de corelaţie a rangurilor a lui *Spearman* se poate deduce şi pe baza coeficientului de corelaţie al lui *Pearson*, notând rangurile cu simbolurile utilizate pentru variantele variabilelor corelate:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n\sigma_x\sigma_y} \tag{18.2}$$

Considerând că rangurile variabilelor X şi Y sunt constituite din primele n numere întregi şi pozitive, nivelul mediu al celor două variabile va fi dat de expresia:

$$\bar{x} = \bar{y} = \frac{\frac{n(n+1)}{2}}{n} = \frac{n+1}{2}$$

iar abaterile medii pătratice se vor calcula astfel:

Scanned with OKEN Scanner

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n} - \bar{x}^2} \quad \text{şi} \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^{n} y_i^2}{n} - \bar{y}^2} \tag{18.3}$$

Suma pătratelor primelor n numere naturale este dată de relaţia:

$$1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{n(n+1)(2n+1)}{6} = \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i^2 \tag{18.4}$$

iar

$$\sigma_x = \sigma_y = \sqrt{\frac{\frac{n(n+1)(2n+1)}{6}}{n} - \left(\frac{n+1}{2}\right)^2} =$$

$$= \sqrt{\frac{n(n+1)(2n+1)}{6n} - \frac{(n+1)^2}{4}} = \sqrt{\frac{2(n+1)(2n+1) - 3(n+1)^2}{12}} = \sqrt{\frac{n^2-1}{12}}$$

$$n\sigma_x\sigma_y = n\frac{n^2-1}{12} \tag{18.5}$$

Notând diferenţa dintre rangurile celor două variabile cu $d_i = x_i - y_i$ sau adunând şi scăzând aceeaşi cantitate $\bar{x} = \bar{y}$, rezultă:

$$d_i = (x_i - \bar{x}) - (y_i - \bar{y})$$

Ridicând la pătrat ambii membri ai acestei relaţii şi însumând, se obţine:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}\left[(x_i - \bar{x}) - (y_i - \bar{y})\right]^2 = \sum_{i=1}^{n}\left[(x_i - \bar{x})^2 - 2(x_i - \bar{x})(y_i - \bar{y}) + (y_i - \bar{y})^2\right] =$$

Scanned with OKEN Scanner

$$= \sum_{i=1}^{n}(x_i - \bar{x})^2 - 2\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) + \sum_{i=1}^{n}(y_i - \bar{y})^2 \qquad (18.6)$$

dar $$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{n(n^2-1)}{12} \qquad (18.7)$$

deci $$\sum_{i=1}^{n} d_i^2 = \frac{2n(n^2-1)}{12} - 2\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \quad .$$

de unde $$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \frac{n(n^2-1)}{12} - \frac{6\sum_{i=1}^{n} d_i^2}{12} \qquad (18.8)$$

Substituind pe (18.5) şi (18.8) în formula coeficientului de corelaţie *Pearson* (18.2), se obţine:

$$r = \frac{\frac{n(n^2-1)}{12} - \frac{6\sum_{i=1}^{n} d_i^2}{12}}{\frac{n(n^2-1)}{12}} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} = 0 \qquad (18.9)$$

La stabilirea rangurilor celor două variabile corelate se poate întâmpla ca două sau mai multe variante să aibă aceeaşi valoare, fapt care nu permite atribuirea unui rang unic şi bine determinat fiecărei variante.

Pentru soluţionarea acestei probleme se poate proceda în două moduri.

Să presupunem că variantele unei variabile ordonate sunt:

5, 6, 7, 7, 7, 8, 10, 12

a) Se poate atribui fiecăreia dintre variantele egale ale unei variabile *rangul celei din centrul grupului* de valori egale:

Scanned with OKEN Scanner

1, 2, 4, 4, 4, 6, 7, 8

b) Se poate atribui fiecăreia dintre variantele egale *rangul pe care îl ocupă prima dintre aceste variante*, astfel:

1, 2, 3, 3, 3, 6, 7, 8

Tabelul 18.1. Distribuţiile unor ţări europene după produsul naţional brut şi după producţia de energie electrică pe un locuitor

Nr. crt.	Ţara	(x_i) $ USA/loc.	(y_i) kwh	Rangul după		d_i	d_i^2
				x_i	y_i		
1	Austria	18800	6590	6	5	1	1
2	Belgia	18490	7190	7	4	3	9
3	Bulgaria	3730	4168	18	12	6	36
4	R.Cehă	7700	5809	14	7	7	49
5	Danemarca	18940	6090	5	6	-1	1
6	Elveţia	23620	8500	1	2	-1	1
7	Franţa	19440	8058	3	3	0	0
8	Germania	20980	5575	2	8	6	36
9	Grecia	8360	3250	13	16	-3	9
10	Italia	18070	3979	8	14	-6	36
11	Norvegia	19130	27429	4	1	3	9
12	Olanda	18050	5085	9	10	-1	1
13	Polonia	5010	3463	17	15	2	4
14	Portugalia	9890	3051	12	18	-6	36
15	Marea Britanie	17750	5534	10	9	1	1
16	România	2910	2426	19	19	0	0
17	R.Slovacă	6450	4308	15	11	4	16
18	Spania	13310	4063	11	13	-2	4
19	Ungaria	6260	3169	16	17	-1	1
	Total	-	-	-	-	-	249

Sursa: Prelucrat după *Anuarul Statistic al României 1995*, CNS, Bucureşti, p.914 şi 919

Remarcăm faptul că, în ambele cazuri, la trecerea de la grupa de variante egale la următoarea unitate statistică, se consideră rangul primei unităţi ce urmează ca

Scanned with OKEN Scanner

şi cum toate elementele precedente ar fi fost neegale. Când numărul perechilor de valori corelate este suficient de mare, nu are importanţă ce metodă se utilizează, dar în cazul unui număr redus de perechi de valori corelate, coeficientul de corelaţie al lui *Spearman* va lua valori diferite.

În tabelul 18.1, după ce au fost atribuite ranguri variantelor celor două variabile s-a determinat diferenţa d_i dintre rangurile variantelor perechi ale variabilelor şi pătratul acestora, pe ultima coloană.

Se observă că nu este necesară ordonarea ţărilor după variantele unei variabile, deoarece diferenţele d_i rămân constante pentru fiecare pereche de variante x_i, y_i indiferent dacă acestea sunt sau nu ordonate.

Aplicând apoi formula coeficientului lui *Spearman* (18.1) rezultă:

$$\theta = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 249}{19^3 - 19} = 0{,}78$$

Deoarece $\theta = 0{,}78$ este mai mare decât zero şi relativ apropiat de +1 rezultă că între nivelul de dezvoltare economică al statelor dat de *PNB/loc.* şi producţia de energie electrică pe un locuitor există o legătură directă şi relativ intensă.

Se remarcă faptul că este indiferent dacă s-au ordonat sau nu unităţile statistice după variaţia variabilei independente, deoarece şi într-un caz şi în celălalt diferenţele dintre ranguri d_i, pentru fiecare unitate statistică, rămân neschimbate.

18.3. Coeficientul de corelaţie a rangurilor al lui *Kendall*

Acest indicator de corelaţie a rangurilor a fost introdus în ştiinţă de către *Fechner* în anul 1897. *M.Kendall* l-a redescoperit în anul 1838 şi l-a studiat din punct de vedere teoretic şi practic ca un indicator neparametric, propunând şi un test specific de verificare a semnificaţiei acestui indicator care de asemeni îi poartă numele.

Fie două variabile statistice X şi Y:

Scanned with OKEN Scanner

$$X: \quad (x_1 \quad x_2 \quad \dots \quad x_i \quad \dots \quad x_n)$$
$$Y: \quad (y_1 \quad y_2 \quad \dots \quad y_i \quad \dots \quad y_n)$$

între care se presupune că există o anumită legătură.

Pentru a obţine un indicator al corelaţiei dintre X şi Y se procedează la ordonarea valorilor variabilei independente X în ordine crescătoare şi alăturat se înscriu variantele corespunzătoare ale variabilei Y. Se determină apoi numărul de valori y_i care nu corespund acestei ordini.

Se notează cu Q numărul de inversiuni ale lui Y faţă de X. Deoarece numărul de perechi distincte x_i, y_i care se pot forma este egal cu $\frac{n(n+1)}{2}$ rezultă că:

$$0 \le Q \le \frac{n(n+1)}{2}$$

Valoarea 0 se obţine atunci când toate variantele y_i se află în aceeaşi ordine ca şi cele ale variabilei X, iar valoarea $\frac{n(n+1)}{2}$ se atinge atunci când valorile y_i se află în ordine total inversă celor ale lui X.

Având în vedere aceste precizări, coeficientul lui *Kendall* notat cu τ poate fi definit prin următoarea formulă:

$$\tau = 1 - \frac{4Q}{n(n-1)} \qquad (18.10)$$

Dacă toate perechile de valori x_i, y_i se află în ordine crescătoare, atunci $Q = 0$ iar

$$\tau = 1 - \frac{4 \cdot 0}{n(n-1)} = 1$$

Dacă toate perechile de valori x_i, y_i se află în ordine inversă , atunci

$$Q = \frac{n(n-1)}{2}$$

Scanned with OKEN Scanner

iar

$$\tau = 1 - \frac{4\dfrac{n(n-1)}{2}}{n(n-1)} = -1$$

Deci $$-1 \le \tau \le +1 .$$

O altă expresie a acestui coeficient de corelaţie, bazat pe rangurile variabilelor corelate, a fost propusă de statisticianul *M.G.Kendall*:

$$\tau = \frac{S}{\frac{1}{2}n(n-1)} \qquad (18.11)$$

în care $S = P - Q$ $P = \sum_{i=1}^{n} p_i$ $Q = \sum_{i=1}^{n} q_i$

Formulele (18.10) şi (18.11) sunt identice, dacă nu există nici o pereche de valori $x_i = x_j$ sau $y_i = y_j$, cazuri în care nu există nici o corespondenţă şi nici discordanţă de ordine $p_i = 0$ şi $q_i = 0$ şi nu se iau în calcul.

În vederea aplicării acestei formule, se procedează la ordonarea variantelor variabilei independente şi alăturat se înscriu nivelurile corespunzătoare ale variabilei dependente. Pe următoarele două coloane se înscriu rangurile celor două variabile. Cu p_i s-a notat numărul de ranguri ale variabilei dependente, superioare rangului i , care există pe coloana respectivă după fiecare rang; iar q_i reprezintă numărul de ranguri inferioare ale variabilei dependente, care există după fiecare rang al acesteia.

Coeficientul (18.11) lui *Kendall* variază tot între -1 şi +1. Astfel, dacă între rangurile celor două variabile există o concordanţă deplină, atunci $Q = 0$, şi $\tau = 1$, iar dacă între ranguri există o discordanţă totală , atunci $P = 0$ şi $\tau = -1$.

Pentru exemplificarea calculului şi analizei coeficientului lui *Kendall*, considerăm exemplul anterior (tabelul 18.1). În acest scop, ţările vor fi ordonate crescător după rangurile variabilei independente şi alăturat se vor înscrie rangurile corespunzătoare ale variabilei dependente (tabelul 18.2). Apoi, pe ultimele două coloane, se vor determina variantele variabilelor p_i şi q_i.

Scanned with OKEN Scanner

Tabelul 18.2. Calculele intermediare ale coeficientului lui *Kendall*

Nr. crt.	Ţara	Rangul după		p_i	q_i
		x	y		
1	Elveţia	1	2	17	1
2	Germania	2	8	11	5
3	Franţa	3	3	15	1
4	Norvegia	4	1	15	0
5	Danemarca	5	6	13	2
6	Austria	6	5	13	1
7	Belgia	7	4	13	0
8	Italia	8	14	5	6
9	Olanda	9	10	8	2
10	Regatul Unit	10	9	8	1
11	Spania	11	13	5	3
12	Portugalia	12	18	1	6
13	Grecia	13	16	2	4
14	R. Cehă	14	7	5	0
15	R. Slovacă	15	11	4	0
16	Ungaria	16	17	1	2
17	Polonia	17	15	1	1
18	Bulgaria	18	12	1	0
19	România	19	19	0	0
	Total	-	-	138	35

Înlocuind în formula (18.11), rezultă:

$$\tau = \frac{2(138-35)}{19^2-19} = 0{,}60$$

$\tau = 0{,}60$ înseamnă că legătura dintre gradul de dezvoltare economică a ţărilor şi nivelul producţiei de energie electrică pe un locuitor este directă şi relativ intensă.

Se observă că valoarea lui τ este mai mică decât cea a lui θ.

În cazul unor valori suficient de mari ale lui n, *Kendall* a arătat că între coeficienţi de corelaţie a rangurilor θ şi τ are loc următoarea relaţie aproximativă (cu excepţia cazurilor în care cei doi coeficienţi se apropie ca valoare de unitate):

$$\theta \approx \frac{3\tau}{2} \qquad (18.12)$$

Scanned with OKEN Scanner

Faţă de θ, calculul coeficientului de corelaţie τ al lui *Kendall* este mai dificil, dar prezintă o serie de avantaje:

a) se poate verifica semnificaţia lui τ prin testele obişnuite în statistică, în timp ce pentru θ, aceasta se poate face numai într-un caz particular, când n este suficient de mare;

b) coeficientul lui *Kendall* poate fi utilizat şi pentru studiul corelaţiei parţiale a rangurilor.

Pentru compararea intensităţii legăturii dintre diferite fenomene este necesar să se utilizeze acelaşi indicator al intensităţii corelaţiei. Este cu totul greşit să se utilizeze pentru comparaţie indicatori diferiţi de corelaţie, de exemplu, coeficientul lui *Pearson* şi un coeficient de corelaţie a rangurilor, deoarece între aceşti indicatori există importante deosebiri metodologice. Mai mult decât atât, coeficienţii de corelaţie simplă, trataţi în acest capitol, vor avea valori diferite, chiar dacă se calculează pentru aceleaşi perechi de valori corelate.

18.4. Coeficientul de concordanţă al lui *Fechner*

Statisticianul german *G.T.Fechner* a stabilit un indicator al intensităţii corelaţiei, calculat fie pe baza concordanţelor şi discordanţelor de semne ale diferenţelor $\Delta x_i, \Delta y_i$ dintre termenii consecutivi ai fiecăreia dintre cele două variabile corelate, fie cu ajutorul concordanţelor şi respectiv, discordanţelor de semn ale abaterilor calculate faţă de mediile variabilelor corelate.

Coeficientul de concordanţă propus de *Fechner*, cunoscut şi sub denumirea de *coeficient simplu de covariaţie diferenţială*, se calculează cu ajutorul formulei:

$$k = \frac{c-d}{c+d} \quad \text{sau} \quad k = \frac{c-d}{n} \quad \text{unde} \quad c+d=n \qquad (18.13)$$

în care c şi d reprezintă numărul de concordanţe, respectiv de discordanţe, de semne ale abaterilor şi n, numărul perechilor de valori corelate:

$$\Delta_{x_i} = x_i - x_{i-1} \quad \text{şi} \quad \Delta_{y_i} = y_i - y_{i-1}$$

sau

$$\Delta'_{x_i} = x_i - \bar{x} \quad \text{şi} \quad \Delta'_{y_i} = y_i - \bar{y}$$

Scanned with OKEN Scanner

Dacă una sau mai multe dintre diferenţele Δ_{x_i} sau Δ_{y_i} este nulă, atunci, perechile respective x_i, y_i ale celor două variabile sunt excluse din calcul.

Din formula (18.13), rezultă că acest indicator variază între -1 şi $+1$:

$$-1 \le k \le +1$$

Astfel, dacă între Δ_{x_i} şi Δ_{y_i} există numai concordanţe de semne, înseamnă că $c = n$ şi $d = 0$ iar $k = 1$; dacă între Δ_{x_i} şi Δ_{y_i} există numai discordanţe de semne, atunci $d = n$ şi $c = 0$, iar coeficientul de concordanţă $k = -1$. Coeficientul de concordanţă al lui *Fechner* este nul, în cazul în care numărul concordanţelor este egal cu numărul discordanţelor $c = d$ şi deci, $k = 0$.

Coeficientul simplu de concordanţă k prezintă dezavantajul că nu ia în calcul mărimea abaterilor Δ_{x_i} şi Δ_{y_i}, ci ţine seama numai de semnele acestor abateri. Pentru a se înlătura această deficienţă, se calculează aşa numitul *coeficient ponderat de concordanţă* după formula:

$$K = \frac{C - D}{C + D} \tag{18.14}$$

unde C - reprezintă suma produselor $\Delta_{x_i}\Delta_{y_i}$ pozitive,

D - reprezintă valoarea absolută a sumei produselor $\Delta_{x_i}\Delta_{y_i}$ negative.

La fel ca şi k, coeficientul K variază între -1 şi $+1$ şi are aceeaşi semnificaţie. Dacă $D = 0$, atunci $K = 1$ şi dacă $C = 0$ atunci $K = -1$, iar când $C = D$ coeficientul ponderat de concordanţă este nul.

Atât coeficientul simplu de concordanţă cât şi cel ponderat au dezavantajul că ei pot lua valorile ± 1, chiar atunci când legătura dintre variabile nu este funcţională, adică diferenţele Δ_{x_i} şi Δ_{y_i} nu sunt proporţionale. Acest neajuns se înlătură dacă se utilizează o altă variantă a coeficientului ponderat de concordanţă, mai apropiată ca structură de coeficientul de corelaţie al lui *Bravais - Pearson*:

$$K' = \frac{\sum_{i=1}^{n} \Delta_{x_i}\Delta_{y_i}}{\sqrt{\sum_{i=1}^{n}(\Delta_{x_i})^2 \sum_{i=1}^{n}(\Delta_{y_i})^2}} \tag{18.15}$$

Scanned with OKEN Scanner

Acest coeficient ia valoarea ± 1 numai în cazul în care diferenţele Δ_{y_i} variază direct proporţional cu Δ_{x_i}.

Pentru exemplificarea metodei de calcul şi analiză a acestor indicatori, considerăm din nou exemplul din tabelul 18.1. Calculele intermediare se prezintă în tabelul 18.3.

$$c = 14 \qquad d = 5$$

Coeficientul simplu de concordanţă:

$$k = \frac{c-d}{n} = \frac{14-5}{19} = \frac{9}{19} = 0{,}47$$

1

Tabelul 18.3.

Nr. crt.	Ţara	mii $ SUA/ loc.	mii Kwh/ loc.	Δ_{x_i}	Δ_{y_i}	c sau d	$\Delta_{x_i}\Delta_{y_i}$	$\Delta^2_{x_i}$	$\Delta^2_{y_i}$
1	Austria	18,8	6,6	5,3	0,4	*c*	2,12	28,09	0,16
2	Belgia	18,5	7,2	5,0	1,0	*c*	5,00	25,00	1,00
3	Bulgaria	3,7	4,2	- 9,8	-2,0	*c*	19,04	96,04	4,00
4	R.Cehă	7,7	5,8	- 5,8	-0,4	*c*	2,32	33,64	0,16
5	Danemarca	18,9	6,1	5,4	-0,1	*d*	0,54	29,16	0,01
6	Elveţia	23,6	8,5	10,1	2,3	*c*	23,23	102,01	5,29
7	Franţa	19,4	8,1	5,9	1,9	*c*	11,21	34,81	3,61
8	Germania	21,0	5,6	7,5	-0,6	*d*	- 4,50	56,25	0,36
9	Grecia	8,4	3,3	- 5,1	-2,9	*c*	14,79	26,01	8,41
10	Italia	18,1	4,0	4,6	-2,2	*d*	10,12	21,16	4,84
11	Norvegia	19,1	27,4	5,6	21,2	*c*	118,72	31,36	449,44
12	Olanda	18,1	5,1	4,6	-1,1	*d*	- 5,06	21,16	1,21
13	Polonia	5,0	3,5	- 8,5	-2,7	*c*	22,95	72,25	7,29
14	Portugalia	9,9	3,1	- 3,6	-3,1	*c*	11,16	12,96	9,61
15	Regatul Unit	17,8	5,5	4,3	-0,7	*d*	-3,01	18,49	0,49
16	România	2,9	2,4	10,6	-3,8	*c*	40,28	112,36	14,44
17	R. Slovacă	6,5	4,3	- 7,0	-1,9	*c*	13,30	49,00	3,61
18	Spania	13,3	4,1	- 0,2	-2,1	*c*	0,42	0,04	4,41
19	Ungaria	6,3	3,2	- 7,2	-3,0	*c*	21,16	51,84	9,00
Total		13,5	6,2	--	--	--	--	821,63	527,34

Coeficientul ponderat de concordanţă:

Scanned with OKEN Scanner

$$K = \frac{C - D}{C + D} = \frac{306,7 - 23,23}{329,93} = 0,86$$

Coeficientul pearsonian de concordanţă:

$$K' = \frac{\sum_{i=1}^{n} \Delta_{x_i} \Delta_{y_i}}{\sqrt{\sum_{i=1}^{n} (\Delta_{x_i})^2 \sum_{i=1}^{n} (\Delta_{y_i})^2}} = \frac{283,47}{\sqrt{821,63 \cdot 527,34}} = \frac{283,47}{658,24} = 0,43$$

Toţi aceşti trei coeficienţi de concordanţă arată că între gradul de dezvoltare economică şi nivelul producţiei de energie electrică produsă există o legătură directă şi relativ intensă.

Valorile acestor coeficienţi sunt diferite datorită diferenţelor metodologice dintre ei. Desigur că dintre aceştia vom acorda mai multă încredere coeficientului pearsonian de concordanţă, care prezintă cele mai puţine deficienţe metodologice.

Scanned with OKEN Scanner

19. ASOCIAŢIA ÎNTRE CARACTERISTICI CALITATIVE

19.1. Noţiunea de asociaţie

Indicatorii parametrici de regresie şi corelaţie nu pot fi utilizaţi decât pentru caracterizarea legăturii dintre două sau mai multe variabile (caracteristici) exprimate numeric. Dacă variabilele Y, a căror legătură se studiază, sunt de esenţă calitativă şi nu pot fi exprimate numeric (culori, categorii sociale, profesii, ocupaţii, opinii etc), atunci legătura dintre ele se poate studia cu ajutorul *metodei asociaţiei*. Informaţiile iniţiale cu privire la variabilele corelate se sistematizează sub forma unui *tabel cu dublă intrare* în care unităţile de observare se grupează concomitent după ambele caracteristici studiate. Acest tabel poartă denumirea de *tabel de contingenţă* sau *tabel de asociaţie* $k \cdot h$ (tabelul 19.1).

În general, studiul asociaţiei dintre fenomene este îndreptat în două direcţii principale:

- studiul existenţei asociaţiei;
- măsurarea intensităţii asociaţiei.

Tabelul 19.1. Schema unui tabel de contigenţă $k \cdot h$

Y \ X	x_1	x_2	...	x_i	...	x_k	Total
y_1	n_{11}	n_{21}	...	n_{i1}	...	n_{k1}	$n_{.1}$
y_2	n_{12}	n_{22}	...	n_{i2}	...	n_{k2}	$n_{.2}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
y_j	n_{1j}	n_{2j}	...	n_{ij}	...	n_{kj}	$n_{.j}$
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
y_h	n_{1h}	n_{2h}	...	n_{ih}	...	n_{kh}	$n_{.h}$
Total	$n_{1.}$	$n_{2.}$	...	$n_{i.}$	...	$n_{k.}$	$n_{..}$

Scanned with OKEN Scanner

19.2. Existenţa asociaţiei

O caracteristică X este independentă de Y, dacă proporţia unităţilor care sunt x_i şi y_j, printre toate cele care sunt y_j este aceeaşi ca şi proporţia celor care sunt x_i, faţă de colectivitatea totală (tabelul $k \cdot h$), pentru orice valoare a lui i şi j, astfel:

$$\frac{n_{ij}}{n_{.j}} = \frac{n_{i.}}{n} \qquad \begin{matrix} i = \overline{1,k} \\ j = \overline{1,h} \end{matrix} \tag{19.1}$$

de unde

$$n_{ij} = \frac{n_{i.} n_{.j}}{n} \tag{19.2}$$

Această formulă arată că, dacă variabila X este independentă de Y, atunci şi Y este independentă de X. Deci, dacă X şi Y sunt independente, atunci frecvenţa relativă din fiecare dreptunghi i, j trebuie să fie egală cu produsul frecvenţelor relative marginale corespunzătoare, adică:

$$\frac{n'_{ij}}{n} = \frac{n_{i.}}{n} \cdot \frac{n_{.j}}{n} \tag{19.3}$$

de unde

$$n'_{ij} = \frac{n_{i.} n_{.j}}{n} \tag{19.4}$$

Abaterile dintre frecvenţele empirice şi frecvenţele teoretice $n_{ij} - n'_{ij}$ măsoară dezacordul dintre caracteristicile studiate.

În cazul unui tabel de contigenţă $2 \cdot 2$, dacă notăm cu a, b, c, d frecvenţele empirice şi cu $\alpha, \beta, \gamma, \delta$ frecvenţele teoretice corespunzătoare, se obţine tabelul 19.2.

În statistică, pentru studierea existenţei asociaţiei, există două metode mai des utilizate:

Scanned with OKEN Scanner

Metoda 1. Se verifică ipoteza de independenţă, adică dacă $p_1 = \frac{a}{n_1}$ este semnificativ diferit de $p = \frac{m_1}{n}$ sau dacă diferenţa $(a - \alpha)$ este semnificativ diferită de zero, în care: $\alpha = \frac{n_1 m_1}{n}$

Tabelul 19.2. Tabel de asociaţie $2*2$ cu frecvenţe teoretice şi empirice

X / Y	x_1	x_2	Total
y_1	$a^{(\alpha)}$	$b^{(\beta)}$	m_1
y_2	$c^{(\gamma)}$	$d^{(\delta)}$	m_2
Total	n_1	n_2	n

Se poate lua în considerare oricare dintre cele patru diferenţe între frecvenţele teoretice şi cele empirice, pentru a caracteriza dezacordul dintre ele, deoarece:

$$|a - \alpha| = |b - \beta| = |c - \gamma| = |d - \delta| \qquad (19.5)$$

Pentru verificarea existenţei asociaţiei între două variabile prezentate într-un tabel de asociaţie $2*2$, se poate utiliza testul *t*, dat de formula:

$$t = \frac{a - \alpha}{\sigma} \qquad (19.6)$$

unde abaterea standard σ a diferenţei $(a - \alpha)$ 1 este definită de relaţia:

$$\sigma^2 = n_1 pq(1 - \frac{n_1}{n}) = n_1 \frac{m_1}{n}(1 - \frac{m_1}{n})(1 - \frac{n_1}{n}) = \frac{n_1 n_2 m_1 m_2}{n^3} \qquad (19.7)$$

Se determină $P(t)$ pe baza tabelelor :

Scanned with OKEN Scanner

$$P(t) = \frac{1}{\sqrt{2\pi}} \int_{t}^{\infty} e^{\frac{-t^2}{2}} dt \qquad (19.8)$$

în funcţie de care se acceptă sau se respinge ipoteza.

Metoda 2. Se caracterizează dezacordul dintre cele două feluri de frecvenţe (teoretice şi empirice), cu ajutorul criteriului χ^2:

$$\chi^2 = \frac{(a-\alpha)^2}{\alpha} + \frac{(b-\beta)^2}{\beta} + \frac{(c-\gamma)^2}{\gamma} + \frac{(d-\delta)^2}{\delta} \qquad (19.9)$$

Se poate demonstra că $t^2 = \chi^2$, unde χ^2 are un singur grad de libertate.

Pe baza acestei identităţi, expresia lui χ^2 devine:

$$\chi^2 = \frac{(a-\alpha)^2}{\sigma^2} = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(a+c)(d+b)(d+c)} = \frac{n(ad-bc)^2}{n_1 n_2 m_1 m_2} \qquad (19.10)$$

Relaţia (19.10) arată că χ^2 este egal cu zero, atunci când produsul de pe diagonala $a \cdot d$ este egal cu produsul $b \cdot c$ de pe cealaltă diagonală. Acest fapt este important deoarece el va indica dacă este necesar sau nu să se continue testul de semnificaţie. Astfel, dacă cele două produse sunt egale sau apropiate ca mărime, atunci χ^2 va fi prea mic pentru a valida ipoteza nulă. Produsele de pe diagonalele tabelului pot fi folosite, de asemenea, pentru a determina direcţia legăturii de asociaţie, în sensul că produsul mai mare ne va indica concentrarea mai mare a valorilor şi deci, direcţia legăturii.

Să considerăm pentru exemplificare distribuţia populaţiei lumii pe mediile urban şi rural şi după nivelul de dezvoltare economică în anul 1990 şi să verificăm dacă există vreo legătură între aceste două caracteristici.

Scanned with OKEN Scanner

Tabelul 19.3. Distribuţia populaţiei lumii după nivelul de dezvoltare economică şi pe medii sociale (urban şi rural) în anul 1990

Gradul de dezvoltare / Mediul social	Ţări dezvoltate	Ţări slab dezvoltate	Total
Urban	881	1401	2282
Rural	330	2683	3013
Total	1211	4084	5295

Sursa: United Nations World Urbanization Prospects 1990, New York, 1991, pp.8-9

Dacă n-ar exista o asociere (legatură) între nivelul de dezvoltare economică şi repartiţia populaţiei pe mediile urban şi rural, atunci ar trebui ca populaţia lumii să fie distribuită între lumea dezvoltată şi cea slab dezvoltată în aceeaşi proporţie ca şi pe mediile urban şi rural, adică frecvenţele teoretice α, β, γ şi δ să fie egale cu cele empirice a, b, c şi d sau ca diferenţele dintre aceste două categorii de frecvenţe să fie nesemnificative.

Pentru aceasta, este necesar un criteriu obiectiv de decizie şi acesta este criteriul χ^2.

Pentru aplicarea acestui criteriu este necesar să se determine, în prealabil, frecvenţele teoretice n_i', pe baza formulei (19.4), astfel:

$$\alpha = \frac{n_1 m_1}{n} = \frac{1211 \cdot 2282}{5295} = 521{,}91$$

$$\beta = \frac{n_2 m_1}{n} = \frac{4084 \cdot 2282}{5295} = 1760{,}1$$

$$\gamma = \frac{n_1 m_2}{n} = \frac{1211 \cdot 3013}{5295} = 689{,}1$$

$$\delta = \frac{n_2 m_2}{n} = \frac{4084 \cdot 3013}{5295} = 2323{,}91$$

Scanned with OKEN Scanner

Verificarea existenţei legăturii dintre nivelul de dezvoltare economică şi repartiţia pe mediile urban şi rural se va efectua prin aplicarea criteriului χ^2. În acest scop, calculele se sistematizează sub forma tabelului 19.4.

Tabelul 19.4. Calculul lui χ^2

	Frecvenţele empirice n_i	Frecvenţele teoretice n_i'	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
A	881	521,91	359,09	128945,63	247,06
B	1401	1760,10	-359,0	128945,63	73,26
C	330	689,10	-359,10	128945,63	187,12
D	2683	2323,91	359,09	128945,63	55,48
Total	5295	5295,02	0	-	562,92

Frecvenţele teoretice se mai pot obţine şi pe altă cale ceva mai simplă.

Pornind de la expresia lui χ^2 şi efectuând calculele cerute de semne, rezultă:

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - n_i')^2}{n_i'} = \sum_{i=1}^{n} \frac{n_i^2 - 2n_i n_i' + n_i'^2}{n_i'} = \sum_{i=1}^{n} \frac{n_i^2}{n_i'} - 2\sum_{i=1}^{n} n_i + \sum_{i=1}^{n} n_i' \qquad (19.11)$$

Ţinând seama de relaţia $\sum_{i=1}^{n} n_i = \sum_{i=1}^{n} n_i'$, rezultă:

$$\chi^2 = \sum_{i=1}^{n} \frac{n_i^2}{n_i'} - n \qquad (19.12)$$

Tabelul 19.5. Calculul lui χ^2 pe baza formulei 19.12

	n_i^2	$\frac{n_i^2}{n_i'}$
a	776161	1487,15
b	1962801	1115,16
c	108900	158,03
d	7198489	3097,58
Total	-	5857,92

Scanned with OKEN Scanner

Pentru determinarea lui χ^2 după formula (19.12), calculul se poate sistematiza cu ajutorul tabelului 19.5.

Aplicând formula 19.12 rezultă:

$$\chi^2 = \sum_{i=1}^{n} \frac{n_i^2}{n_i'} - n = 5857{,}92 - 5295 = 562{,}92$$

Calculul lui χ^2, în cazul unui tabel de contingenţă $2 \cdot 2$, se poate face şi pe baza formulei (19.10):

$$\chi^2 = \frac{n(ad-bc)^2}{n_1 n_2 m_1 m_2} = \frac{5295(881 \cdot 2683 - 1401 \cdot 330)^2}{1211 \cdot 4084 \cdot 2282 \cdot 3013} = 562{,}92$$

Se observă că toate cele trei metode de determinare a lui χ^2, în cazul unui tabel $2 \cdot 2$, dau acelaşi rezultat.

În cazul general al unui tabel de contingenţă $k \cdot h$, calculul lui χ^2 se face după formula:

$$\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{(n_{ij} - n_{ij}')^2}{n_{ij}'} \qquad (19.13)$$

în care $$n_{ij}' = \frac{n_{i.} n_{.j}}{n}$$

Efectuând calculele în formula (19.13), rezultă:

$$\chi^2 = n\left(\sum_{i=1}^{k} \sum_{j=1}^{h} \frac{n_{ij}^2}{n_{i.} n_{.j}} - 1\right) \qquad (19.14)$$

Din formula (19.14) se observă că, pentru determinarea lui χ^2, nu mai este necesar să se calculeze în prealabil frecvenţele teoretice n_{ij}'.

O problemă importantă ce trebuie rezolvată pentru aplicarea testului χ^2 este determinarea numărului *gradelor de libertate*.

Scanned with OKEN Scanner

În cazul general al unui tabel de contigenţă $k \cdot h$, numărul gradelor de libertate f va fi dat de formula:

$$f = (k-1)(h-1) \tag{19.15}$$

Este clar că în cazul unui tabel de contigenţă $2 \cdot 2$ numărul gradelor de libertate va fi egal cu:

$$f = (2-1)(2-1) = 1 \tag{19.16}$$

iar în cazul unui tabel de contigenţă $3 \cdot 3$:

$$f = (3-1)(3-1) = 4 \tag{19.17}$$

Pe baza numărului gradelor de libertate şi la un anumit nivel de semnificaţie, se determină, din tabele speciale, valoarea teoretică a lui χ^2. Dacă valoarea lui χ^2, obţinută prin calcule, este mai mare decît valoarea tabelară, înseamnă că ipoteza nulă se respinge şi invers.

Astfel, în exemplul anterior (tabelul 19.3), la un nivel de semnificaţie $q_i = 0{,}001$ se găseşte o valoare $\chi_{1;0,999} = 10{,}8$. Deci valoarea lui χ^2 obţinută prin calcule 3 este mai mare decât valoarea tabelară, ceea ce înseamnă că ipoteza nulă se respinge, cu un risc de a greşi într-un caz din 100. Aceasta înseamnă că există o legătură reală între gradul de dezvoltare economică a statelor lumii şi gradul lor de urbanizare , în sensul că lumea mai dezvoltată din punct de vedere economic are şi un grad mai ridicat de urbanizare şi invers.

Pentru verificarea existenţei asociaţiei, în cazul general al unui tabel $k \cdot h$, să considerăm distribuţia căsătoriilor după starea civilă a soţilor înainte de căsătorie în anul 1994 (tabelul. 19.6).

Tabelul 19.6. Căsătorii după starea civilă a soţilor înainte de căsătorie (mii)

Starea civilă a soţului \ Starea civilă a soţiei	Necăsătorită	Văduvă	Divorţată şi nedeclarat	Total
Necăsătorit	125,4	0,8	6,7	132,9
Văduv	0,6	1,2	1,1	2,9
Divorţat şi nedeclarat	8,9	1,3	8,2	18,4
Total	134,9	3,3	16,0	154,2

Sursa: Anuarul statistic al României, 1995, CNS, Bucureşti, pp.120-121

Scanned with OKEN Scanner

Pentru verificarea existenţei unei legături între starea civilă a soţilor înainte de casătorie, vom proceda la verificarea ipotezei nule cu ajutorul testului χ^2.

Ipoteza nulă presupune că între starea civilă a soţilor înainte de căsătorie nu ar exista nici un fel de legătură.

Considerăm apoi un nivel de semnificaţie egal cu 0,001, iar numărul gradelor de libertate va fi egal cu: $f = (3-1)(3-1) = 4$.

Ca şi în cazul anterior, calculul lui χ^2, în baza formulei (19.14), se va sistematiza cu ajutorul tabelului 19.4.

Tabelul 19.4. Calculul lui χ^2 în cazul unui tabel de contingenţă $3 \cdot 3$

Celula	n_{ij}	n'_{ij}	n_{ij}^2	$\frac{n_{ij}^2}{n'_{ij}}$
n_{11}	125,39	116,28	15722,65	135,21
n_{12}	0,79	2,84	0,62	0,22
n_{13}	6,73	13,78	45,29	3,29
n_{12}	0,61	2,54	0,37	0,15
n_{22}	1,23	0,06	1,51	25,17
n_{23}	1,06	0,30	1,12	3,73
n_{31}	8,93	16,10	79,74	4,95
n_{32}	1,28	0,39	1,64	4,21
n_{33}	8,20	1,91	67,24	35,20
Total	154,22	154,20	-	212,13

$$\chi^2_{calc} = \sum_{i=1}^{k}\sum_{j=1}^{h}\frac{n_{ij}^2}{n'_{ij}} - n = 212{,}13 - 154{,}22 = 57{,}91$$

$$\chi^2_{tab} = \chi^2_{4;99{,}9} = 18{,}5$$

Deci, $\chi^2_{calc} > \chi^2_{tab}$ înseamnă că ipoteza nulă se respinge, deci există o legătură între starea civilă a soţilor înainte de căsătorie, în sensul că se căsătoresc de regulă persoane cu aceeaşi stare civilă avută înainte de căsătorie.

Scanned with OKEN Scanner

19.3. Intensitatea asociaţiei

În continuare se vor prezenta unii indicatori ai intensităţii asociaţiei cu ajutorul cărora se poate caracteriza numeric intensitatea legăturii dintre variabilele exprimate prin cuvinte, ca şi în cazul caracteristicilor cantitative .

În cazul unui tabel de contigenţă $2 \cdot 2$ (tabelul 19.2.), se pot ivi două situaţii opuse:

• *independenţă totală*, când are loc relaţia:

$$ad = bc \tag{19.18}$$

• *asociaţie pozitivă completă* când $bc = 0$. Aceasta se poate întâmpla în una dintre următoarele situaţii:

$$\begin{array}{l} 1)\ b = 0 \text{ şi } c \neq 0 \\ 2)\ c = 0 \text{ şi } b \neq 0 \\ 3)\ b = 0 \text{ şi } c = 0 \end{array} \tag{19.19}$$

• *asociaţia negativă completă* , când $ad = 0$, adică:

$$\begin{array}{l} 1)\ a = 0 \text{ şi } d \neq 0 \\ 2)\ d = 0 \text{ şi } a \neq 0 \\ 3)\ a = 0 \text{ şi } d = 0 \end{array} \tag{19.20}$$

Dintre aceste situaţii notăm faptul că, în cea de-a treia variantă, asociaţia este mai completă decât în primele două, atât în cazul asociaţiei pozitive, cât şi în cazul celei negative.

În literatura de specialitate se cunosc numeroşi indicatori ai intensităţii asociaţiei dintre variabile, dar aici vor fi prezentaţi numai cei mai reprezentativi.

Coeficientul de asociaţie al lui G.U.Yulle este definit de relaţia:

$$Q = \frac{ad - bc}{ad + bc} \tag{19.21}$$

Din formula (19.21), rezultă: $-1 \leq Q \leq 11$

Astfel: $Q = 0$, dacă $ad = bc$, deci în cazul când variabilele luate în studiu sunt independente;

Scanned with OKEN Scanner

$Q = 0$, dacă $bc = 0$, adică între variabile există o asociaţie pozitivă completă;

$Q = -1$, dacă $ad = 0$, când între variabile există o asociaţie negativă completă.

Acest indicator prezintă dezavantajul că nu face nici o distincţie între cele trei cazuri posibile de asociaţie completă pozitivă sau negativă.

Considerând pentru exemplificare datele din tabelul 19.3 şi aplicând formula 19.21, rezultă:

$$Q = \frac{ad - bc}{ad + bc} = \frac{881 \cdot 2683 - 1401 \cdot 330}{881 \cdot 2683 + 1401 \cdot 330} = 0{,}67$$

$Q = 0{,}67$ arată că între nivelul de dezvoltare economică a lumii şi gradul de urbanizare există o legătură directă şi relativ intensă.

Un alt indicator de măsurare a intensităţii asociaţiei, în cazul unui tabel de asociaţie $2 \cdot 2$, este *coeficientul de coligaţie al lui Yulle*, care are aceleaşi proprietăţi şi deci, şi aceleaşi deficienţe ca şi coeficientul său de asociaţie.

Coeficientul de coligaţie Y este dat de relaţia:

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} \qquad (19.22)$$

Aplicând formula 19.22 la exemplul din tabelul 19.3, se obţine:

$$Y = \frac{\sqrt{ad} - \sqrt{bc}}{\sqrt{ad} + \sqrt{bc}} = \frac{\sqrt{881 \cdot 2683} - \sqrt{1401 \cdot 330}}{\sqrt{881 \cdot 2683} + \sqrt{1401 \cdot 330}} = 0{,}39$$

Aceasta înseamnă că între nivelul de dezvoltare economică şi gradul de urbanizare a lumii există o legătură directă de intensitate medie.

Pătratul de contigenţă ϕ^2 a fost propus de către statisticianul englez *K.Pearson* şi este definit de formula:

$$\phi^2 = \frac{\chi^2}{n} = \frac{(ad - bc)^2}{m_1 n_1 m_2 n_2} \qquad (19.23)$$

sau

Scanned with OKEN Scanner

$$\phi = \frac{ad - bc}{\sqrt{m_1 n_1 m_2 n_2}} \tag{19.24}$$

Din formula (19.24), rezultă că:

$$-1 \le \phi \le 1$$

$\phi = 0$, când $ad = bc$ şi corespunde independenţei dintre cele două caracteristici;

$\phi = \pm 1$, când $a = d = 0$ sau $b = c = 0$

Pătratul de contigenţă este egal cu ± 1, numai în cazul asociaţiilor pozitive şi respectiv, negative complete.

Aplicând formula (19.24) la exemplul din tabelul 19.3, rezultă:

$$\phi = \frac{ad - bc}{\sqrt{m_1 n_1 m_2 n_2}} = \frac{881 \cdot 2683 - 1401 \cdot 330}{\sqrt{2282 \cdot 1211 - 3013 \cdot 4084}} = \frac{1901393}{5831393} = 0{,}33$$

Rezultă că între cele două variabile există o asociaţie pozitivă relativ slabă.

Menţionăm faptul că, pentru aceleaşi date, deci pentru aceeaşi intensitate a asociaţiei, cei trei indicatori Q, Y şi ϕ dau, în mod obişnuit, rezultate diferite. De aceea, pentru compararea intensităţii asociaţiei dintre perechi diferite de variabile, este necesar să se utilizeze acelaşi indicator de intensitate.

În cazul general al tabelelor de contigenţă $k \cdot h$, intensitatea asociaţiei dintre cele două caracteristici se determină cu ajutorul *coeficientului de contigenţă C al lui K.Pearson*, care se bazează pe diferenţele dintre frecvenţele reale şi cele teoretice $n_{ij} - n'_{ij}$ calculate pentru toate perechile i şi j din cadrul tabelului.

Coeficientul de contigenţă este dat de următoarea formulă:

$$C = \frac{\phi^2}{1 + \phi^2} = \frac{\chi^2}{n + \chi^2} \tag{19.25}$$

unde

$$\phi^2 = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{h} \frac{(n_{ij} - n'_{ij})^2}{n'_{ij}} = \frac{\chi^2}{n} \tag{19.26}$$

Scanned with OKEN Scanner

$C = 0$, când $\chi^2 = 0$, iar cele două caracteristici sunt independente. Valoarea lui C creşte, pe măsură ce creşte χ^2. *K.Pearson* a demonstrat că C reprezintă o estimaţie a coeficientului de corelaţie r dintr-o colectivitate.

Un alt indicator al intensităţii îl constituie *coeficientul lui Ciuprov* - notat cu T şi definit de relaţia:

$$T^2 = \frac{\phi^2}{\sqrt{(k-1)(h-1)}} \qquad (19.27)$$

T^2 faţă de C prezintă avantajul că poate atinge valoarea limită superioară, atunci când χ^2 creşte.

Calculând acest indicator pe baza datelor din tabelul 19.6, rezultă:

$$\phi^2 = \frac{57,91}{154,2} = 0,38 \quad \text{iar} \quad T^2 = \frac{0,38}{\sqrt{(3-1)(3-1)}} = 0,19$$

$T^2 = 0,19$ arată că între cele două variabile există o legătură directă de intensitate redusă.

Scanned with OKEN Scanner

20. ELASTICITATEA

20.1. Noţiunea şi felurile ei

Elasticitatea reprezintă o noţiune generală preluată din ştiinţele naturii şi exprimă proprietatea unei funcţii de a-şi modifica valoarea în mod neproporţional, sub influenţa variaţiei argumentului.

În ştiinţele economice, conceptul general de elasticitate apare într-o formă incipientă, în opera fiziocratului *A.R.J.Turgot* (1727-1781), atunci când analizează problema randamentului economic descrescând al pământului (1727). Această idee a fost reluată, un secol mai târziu, de către ultimul discipol al lui *William Petty* , economistul *Gregory King*, în 1896, cu ocazia analizei "expansibilităţii" preţurilor în raport cu cantităţile de grâu recoltate sau ofertate - cunoscută sub denumirea de *legea* sau *efectul King*. *Coeficientul de expansibilitate* a preţurilor al lui *King* se obţine sub forma unui raport între scăderea relativă a preţurilor unitare şi scăderea relativă a cantităţilor de grâu recoltate.

Conceptul economic de elasticitate a fost introdus în ştiinţele economice de către marele economist şi econometrician englez *Alfred Marshall* (1842-1924), reprezentant de seamă al şcolii neoclasice de economie de la *Cambridge*. Pentru ca acest concept să devină operaţional, *Marshall* a elaborat şi o metodă statistică de măsurare a elasticităţii, în anul 1885, cu ajutorul *coeficientului de elasticitate* a cererii, în Cartea a III-a a *Principles of Economics*, publicată în anul 1891 şi intitulată *Nevoile şi satisfacerea lor*. Criticând teoria economică clasică, *Marshall* îmbină, în mod fericit, teoria utilităţii cu cea a cererii într-o expunere marginalistă, matematizată. Astfel, el formulează *legea universală a cererii*: *volumul cererii creşte, atunci când preţurile scad şi scade, atunci când acestea cresc*.

În economie, elasticitatea reprezintă o proprietate a unei variabile economice (cerere, ofertă, consum etc.) de a-şi modifica mărimea, în mod neproporţional, sub influenţa altor variabile (venit, preţ, factori demografici etc.).

Elasticitatea exprimă, deci, *sensibilitatea* sau *flexibilitatea* unei variabile rezultative y, în funcţie de variaţia unei sau unor variabile cauzale (factoriale).

Scanned with OKEN Scanner

De exemplu, mărimea cererii de bunuri şi servicii variază, în general, în funcţie de mărimea veniturilor populaţiei şi variaţia preţurilor, în mod diferit de la un bun sau serviciu la altul sau de la un grup de mărfuri şi servicii la altul. Oferta de mărfuri este sensibilă la variaţia producţiei şi a cererii etc.

În funcţie de numărul de variabile cauzale x_i, $i = \overline{1,n}$, de care depinde variabila rezultativă y, se disting două tipuri principale de elasticitate:

- *elasticitate simplă*, când $i = 1$.
- *elasticitate multiplă*, când $i \geq 2$.

Elasticitatea este o funcţie care poate fi constantă sau variabilă în punctele sale şi poartă numele, în primul caz, de *isoelasticitate* iar în al doilea de *heteroelasticitate*.

Elasticitatea poate fi *punctuală* sau *pe* anumite *intervale*.

În cazul elasticităţii multiple, distingem: elasticitatea funcţiei în raport cu fiecare variabilă independentă, celelalte fiind presupuse constante, numită *elasticitate parţială* şi *elasticitate totală*, determinată în raport cu ansamblul variabilelor independente, fiecare dintre acestea având modelul propriu de variaţie.

Elasticitatea multiplă poate fi analizată în timp şi în spaţiu.

După sensul legăturii dintre y şi x, *elasticitatea* poate fi *directă* sau *inversă*.

20.2. Elasticitatea simplă

Gradul de elasticitate a variabilei y, în funcţie de variaţia variabilei cauzale x, este măsurat cu ajutorul *coeficientului de elasticitate*, notat cu $E_{(y/x)}$ - adică elasticitatea variabilei y în raport cu x - şi se exprimă sub forma unui raport între variaţia relativă (procentuală) a variabilei rezultative $\frac{dy}{y}$ şi variaţia relativă a variabilei factoriale $\frac{dx}{x}$, astfel:

- fie funcţia de elasticitate:

$$y = f(x)$$

atunci coeficientul de elasticitate va avea următoarele trei expresii echivalente care vor duce la formularea a trei definiţii complementare:

$$E_{(y/x)} = \frac{dy}{y} : \frac{dx}{x} \qquad (20.1)$$

Din relaţia (20.1), rezultă că elasticitatea funcţiei $y = f(x)$ este egală cu rata variaţiei relative a lui y, care revine pe o unitate de variaţie relativă a lui x sau elasticitatea funcţiei y în raport cu variabila x este egală cu raportul procentual de variaţie a lui y faţă de variaţia procentuală a lui x.

Din această definiţie rezultă că formula (20.1), ca şi cele echivalente, nu este adevărată decât pentru variaţii infinitezimale:

$$E_{(y/x)} = \lim_{\Delta x \to 0} \left(\frac{\Delta y}{y} : \frac{\Delta x}{x} \right) = \frac{dy}{y} : \frac{dx}{x}$$

Aceasta înseamnă că elasticitatea este egală cu limita raportului dintre variaţia relativă a lui y şi variaţia relativă a lui x, atunci când aceasta din urmă tinde către zero.

Creşterea infinitezimală a unor factori în economie se întâlneşte foarte rar.

Tot din formula (20.1) rezultă că elasticitatea este independentă de unităţile de măsură ale variabilelor x şi y:

$$E_{(y/x)} = \frac{dy}{dx} : \frac{y}{x} \qquad (20.2)$$

Relaţia (20.2) arată că $\frac{dy}{dx}$ reprezintă derivata funcţiei y în raport cu x, ceea ce în economie corespunde valorii marginale a funcţiei $y = f(x)$, iar $\frac{y}{x}$ este valoarea medie unitară a lui y. De aici rezultă a doua definiţie a elasticităţii: *elasticitatea funcţiei y în raport cu variabila x este egală cu raportul între valoarea marginală a funcţiei şi valoarea medie a acesteia.*

Ecuaţia (20.1) este echivalentă cu expresia:

Scanned with OKEN Scanner

$$E_{(y/x)} = \frac{dy}{dx} \cdot \frac{x}{y} \qquad (20.3)$$

Considerând dy şi dx ca diferenţiale:

$$\int \frac{1}{y} dy = lg\, y$$

şi

$$\int \frac{1}{x} dx = lg\, x$$

rezultă

$$\frac{dy}{y} = d(lg\, y) \qquad \text{şi} \qquad \frac{dx}{x} = d(lg\, x)$$

atunci (20.3) devine:

$$E_{(y/x)} = \frac{dy}{y} : \frac{dx}{x} = \frac{d(lg\, y)}{d(lg\, x)} \qquad (20.4)$$

Din (20.4) se poate deduce a treia definiţie a elasticităţii: *elasticitatea funcţiei* y *în raport cu variabila* x *este egală cu raportul între derivata logaritmică a lui* y *şi derivata logaritmică a lui* x.

De aici rezultă şi o altă proprietate importantă a elasticităţii şi anume, coeficientul de elasticitate a unei funcţii într-un punct oarecare al curbei $y = f(x)$ este dat de panta tangentei la curbă, atunci când y şi x se reprezintă grafic pe scări logaritmice.

De exemplu, în cazul unei funcţii liniare $y = a + bx$ care este *heteroelastică* (cu excepţia a trei cazuri particulare de isoelasticitate), elasticitatea va fi:

$$E_{(y/x)} = \frac{dy}{dx} \cdot \frac{x}{y} = \frac{bx}{a + bx}$$

Scanned with OKEN Scanner

Cele trei cazuri particulare de *isoelasticitate* pentru o funcţie liniară sunt:

1) când $a = 0$, deci $y = bx$, funcţia prezintă o elasticitate constantă şi egală cu unitatea.

2) când $x = 0$, deci $y = a$, elasticitatea funcţiei este constant nulă.

3) când $x = constant$, atunci elasticitatea funcţiei este egală cu infinit.

Funcţia putere de forma $y = ax^b$ prezintă o isoelasticitate pentru orice valoare a lui x .

Coeficientul de elasticitate arată cu câte procente se modifică variabila dependentă, atunci când variabila independentă variază cu 1%.

Elasticitatea cererii c_i a unui produs i se poate determina în funcţie de variaţia venitului V, de variaţia preţului produsului i şi de variaţia preţurilor altor produse j (cerere încrucişată), astfel:

$$E_{(c_i/V)} = \frac{dc_i}{c_i} : \frac{dV}{V} \tag{20.5}$$

$$E_{(c_i/p_i)} = \frac{dc_i}{c_i} : \frac{dp_i}{p_i} \tag{20.6}$$

$$E_{(c_i/p_j)} = \frac{dc_i}{c_i} : \frac{dp_j}{p_j} \tag{20.7}$$

Elasticitatea cererii unui produs în funcţie de variaţia preţurilor acestui produs, în perioada curentă t faţă de o perioadă de bază 0 se poate calcula cu ajutorul formulei:

$$E_{(c_i/p_i)} = \frac{\frac{c_{ti} - c_{0i}}{c_{ti} + c_{0i}}}{\frac{2}{2}} : \frac{\frac{p_{ti} - p_{0i}}{p_{ti} + p_{0i}}}{2} = \frac{c_{ti} - c_{0i}}{c_{ti} + c_{0i}} \cdot \frac{p_{ti} - p_{0i}}{p_{ti} + p_{0i}} \tag{20.8}$$

sau sub formă logaritmică:

Scanned with OKEN Scanner

$$E_{(c_t/p_t)} = \frac{lg\, c_t - lg\, c_0}{lg\, p_t - lg\, p_0}$$

1

Coeficientul de elasticitate variază între zero şi infinit:

$$0 \le E_{(y/x)} \le +\infty \qquad (20.9)$$

Pentru ilustrarea modului de interpretare a coeficientului de elasticitate în acest interval, considerăm elasticitatea cererii unui produs în funcţie de preţul unitar al acestuia.

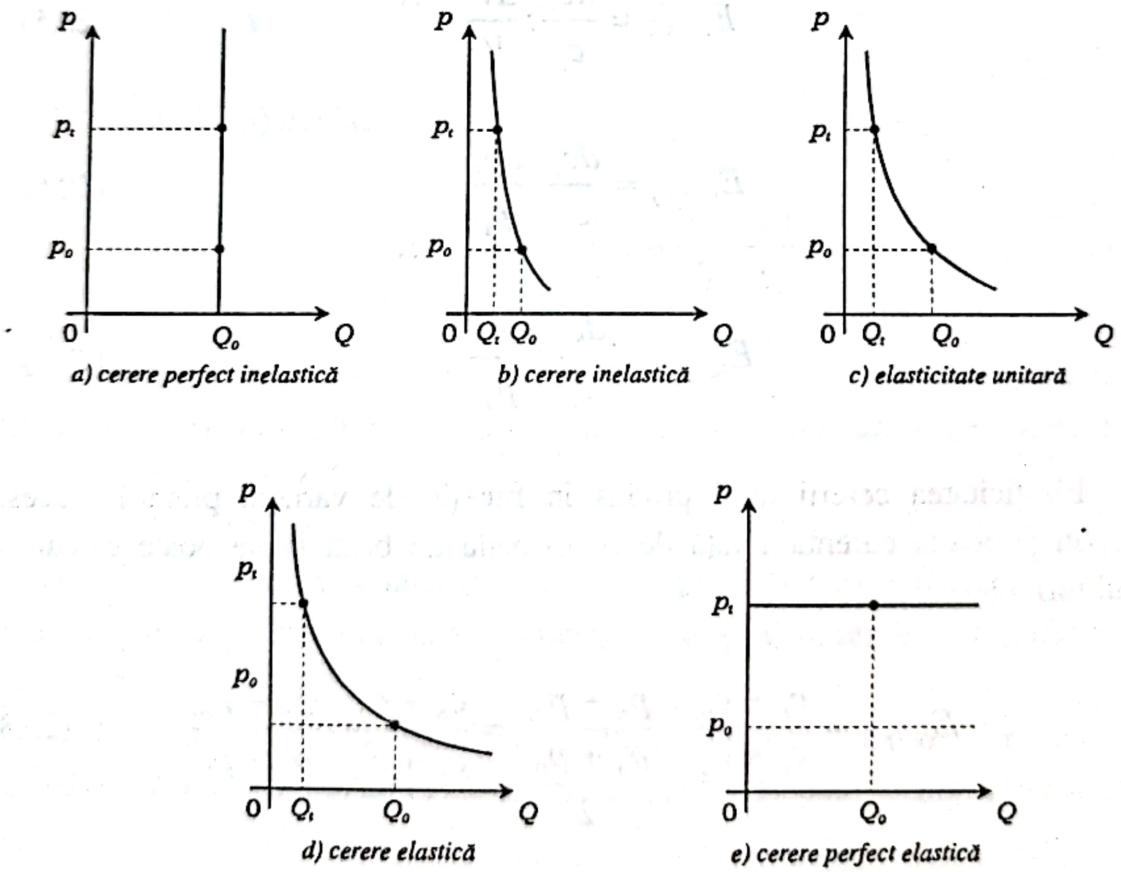


Figura 20.1. Tipuri de elasticităţi

Scanned with OKEN Scanner

Între mărimea cererii unui produs şi preţul lui există, în general, o legătură inversă. Teoretic, însă, se pot distinge următoarele situaţii:

a) *cerere perfect inelastică*, atunci când schimbarea relativă a preţurilor nu exercită nici o influenţă asupra mărimii cererii:

$$\Delta c_i = 0 \qquad E_{(c/p)} = \frac{0}{c} : \frac{\Delta p}{p} = 0 \tag{20.10}$$

b) *cerere inelastică* la produse alimentare de strictă necesitate, atunci când schimbarea relativă a cererii este inferioară schimbării relative a preţurilor:

$$0 < E_{(c/p)} < 1 \tag{20.11}$$

c) *elasticitatea unitară* sau proporţională, atunci când schimbarea relativă a volumului cererii este egală cu schimbarea relativă a preţului:

$$E_{(c/p)} = 1 \tag{20.12}$$

d) *cerere elastică*, întâlnită la produse casnice de uz îndelungat, atunci când schimbarea relativă a volumului cererii este mai mare decât schimbarea relativă a preţului:

$$1 < E_{(c/p)} < \infty \tag{20.13}$$

e) *cerere perfect elastică* apare în cazul produselor de lux, atunci când schimbarea relativă a volumului cererii tinde la infinit, ca urmare a variaţiei relative a preţurilor:

$$E_{(c/p)} = \infty \tag{20.14}$$

Cele cinci tipuri de elasticitate de mai sus se pot ilustra grafic ca în figura 20.1.

Scanned with OKEN Scanner

20.3. Elasticitatea multiplă

Fie o funcţie y de două sau mai multe variabile independente x_i, $i = 2,3,\ldots,n$. Elasticitatea funcţiei în raport cu una dintre variabilele independente x_i, celelalte rămânând constante, se obţine pe baza derivatelor parţiale obişnuite sau a celor logaritmice:

$$y = f(x_i) \qquad i = 2,3,\ldots,n$$

atunci elasticitatea lui y în raport cu x_i va fi:

$$E_{(y/x_2 \cdot x_3 x_4 \ldots x_n)} = \frac{\partial y}{\partial x_2} : \frac{y}{x_2} = \frac{lg\, y}{lg\, x_2}$$

$$E_{(y/x_3 \cdot x_2 x_4 \ldots x_n)} = \frac{\partial y}{\partial x_3} : \frac{y}{x_3} = \frac{lg\, y}{lg\, x_3}$$

.....

$$E_{(y/x_i \cdot x_2 x_3 \ldots x_n)} = \frac{\partial y}{\partial x_i} : \frac{y}{x_i} = \frac{lg\, y}{lg\, x_i} \qquad (20.15)$$

.....

$$E_{(y/x_n \cdot x_2 x_3 \ldots x_{n-1})} = \frac{\partial y}{\partial x_n} : \frac{y}{x_n} = \frac{lg\, y}{lg\, x_n}$$

Ca şi în cazul corelaţiei multiple prin punct s-a separat variabila factorială considerată, în raport cu care se calculează elasticitatea funcţiei, de celelalte variabile independente, care se presupun constante.

Scanned with OKEN Scanner

21. FUNCŢIILE DE PRODUCŢIE

21.1. Noţiunea şi felurile lor

Funcţiile de producţie constituie modele globale ale unui sistem de producţie, reprezentat printr-o singură ecuaţie, care exprimă raportul cantitativ şi stabil între intrări (resurse materiale şi umane) şi ieşiri (rezultate, producţie). Forma generală a unei astfel de funcţii se reprezintă astfel:

$$F(x_i) = 0 \qquad i = \overline{1,n}$$

în care variabilele x_i cu semnul minus reprezintă cheltuielile de resurse, iar cele cu semnul plus, producţia. Resursele sau factorii de producţie se pot exprima în unităţi naturale sau valorice.

Începând din anul 1928, odată cu publicarea de către *Paul H.Douglas* - profesor de economie la Universitatea din Chicago şi *C.W. Cobb* a celebrului studiu *A Theory of Production*[1], funcţiile putere şi-au găsit o largă aplicabilitate în modelarea globală, agregată a fenomenelor economice, atât la nivel micro, cât şi la cel macroeconomic şi au constituit, decenii de-a rândul, un subiect preferat al unui mare număr de cercetători din domeniul economiei, statisticii şi econometriei.

Deoarece prima aplicare în economie a funcţiilor putere a fost modelarea volumului producţiei, ca variabilă rezultativă în funcţie de doi dintre factorii de care aceasta depinde - valoarea capitalului în funcţiune şi cantitatea de forţă de muncă consumată, funcţiile putere sunt cunoscute în ştiinţele economice sub denumirea de *funcţii de producţie*. Sfera largă de aplicabilitate a funcţiilor putere în economie (producţie, productivitatea muncii, costuri, circulaţia mărfurilor, transporturi etc.) în prezent face ca această denumire să fie inadecvată în cazul general, ea se păstrează, însă, datorită popularităţii de care se bucură.

Pornind de la natura informaţiilor statistice cu privire la sistemul de producţie modelat, se pot distinge două categorii de funcţii de producţie: unele, construite pe baza unor indicatori economici de moment (prezentaţi sub formă de stocuri de resurse sau de producţie, numite *funcţii de producţie statice (sincrone)* şi altele,

[1] în *American Economic Review* no.18, pp.139-165

construite pe baza unor serii dinamice, numite *funcţii de producţie dinamice (diacronice).*

Funcţiile de producţie trebuie să satisfacă anumite *condiţii logice, economice şi matematice*, printre care mai importante sunt următoarele:

- toate mărimile care intră în componenţa lor trebuie să fie măsurabile şi să aibă un conţinut economic bine definit;
- producţia nu este posibilă decât prin cheltuirea de resurse;
- resursele sunt necesare şi unele pot fi schimbate între ele în anumite proporţii;
- resursele fiind limitate, producţia nu poate creşte la infinit;
- în general, funcţiile de producţie sunt continue şi derivabile.

O funcţie putere de n variabile este descrisă, în general, de ecuaţia:

$$Y = a(\prod_{i=1}^{n} X_i^{b_i}) + \varepsilon \qquad i = \overline{1,n} \tag{21.1}$$

unde $X_i > 0, Y > o, a > 0$, iar ε este o variabilă aleatoare cu speranţa matematică nulă şi care poate fi neglijată în aplicaţiile practice.

Pentru estimarea parametrilor unui astfel de model de regresie este mai convenabil ca, în prealabil, ecuaţia (21.1) să fie liniarizată prin logaritmare:

$$\lg y_j = \lg a + b_1 \lg x_{1j} + b_2 \lg x_{2j} + \cdots + b_n \lg x_{nj} \qquad j = \overline{1,k}$$

sau mai sintetic:

$$lg\, y_j = lg\, a + \sum_{i=1}^{n} b_i\, lg\, x_{ij} \tag{21.2}$$

Modelul de regresie curbilinie (21.1) se poate particulariza în funcţie de numărul de variabile factoriale x_i luate în calcul, obţinându-se mai multe variante ale funcţiei putere, astfel:

- funcţia putere cu o singură variabilă factorială:

$$Y = aX^b \tag{21.3}$$

- funcţia putere cu două variabile factoriale:

$$Y = aX_1^{b_1} X_2^{b_2} \tag{21.4}$$

Scanned with OKEN Scanner

• funcţia putere cu trei variabile factoriale:

$$Y = aX_1^{b_1}X_2^{b_2}X_3^{b_3} \tag{21.5}$$

• funcţia putere cu n variabile factoriale:

$$Y = a\prod_{i=1}^{n} X_i^{b_i} \tag{21.6}$$

Modelele de regresie curbilinie, descrise de funcţiile putere, îşi găsesc o largă aplicare în teoria şi practica economică şi socială ca modele de trend, ca funcţii de cerere, ca funcţii de producţie etc.

21.2. Funcţia putere cu un singur factor

În cazul luării în calcul a unei singure variabile independente, modelul de regresie simplă descris de funcţia putere (21.3) se prezintă grafic ca în figura 21.1 (A şi B).

În practică, pentru identificarea unei legături dintre variabile de tipul unei funcţii putere, se poate recurge la metoda grafică. Într-un sistem de axe rectangulare cu scări logaritmice pe ambele axe se construieşte corelograma valorilor empirice. Dacă corelograma ia forma unei drepte, înseamnă că legătura dintre variabilele empirice poate fi descrisă de modelul unei funcţii putere cu o singură variabilă factorială.

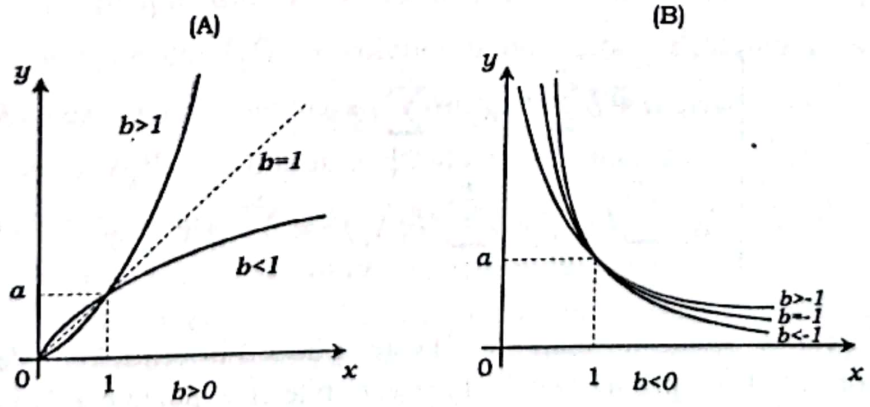


Fig.21.1.Funcţiile de producţie cu un singur factor

Din figura 21.1, rezultă că funcţia putere trece prin punctele de coordonate $(1,a)$, oricare ar fi valorile parametrilor a şi b. În cazul în care b este pozitiv (fig.21.1 A), curba funcţiei putere trece prin originea axelor de coordonate şi este crescătoare şi convexă pentru $b>1$, iar pentru $b<1$, este crescătoare şi concavă şi se transformă într-o dreaptă pentru $b=1$.

Când $b<0$, curbele funcţiei putere trec prin punctul de coordonate $(1,a)$ şi au ca asimptote cele două axe de coordonate (fig.21.1 B).

Modelul de regresie (21.3), liniarizat prin logaritmare, devine:

$$lg\, y_x = lg\, a + b\, lg\, x \tag{21.7}$$

Parametrii ecuaţiei de regresie (21.7) se pot uşor determina prin utilizarea metodei celor mai mici pătrate. Acest fapt se reduce la a determina minimul expresiei:

$$S = \sum_{i=1}^{n} (lg\, y_i - lg\, y_x)^2 = minim$$

sau pe baza ecuaţiei (21.7), se obţine:

$$S = \sum_{i=1}^{n} (lg\, y_i - lg\, a - b\, lg\, x)^2 = minim$$

Calculând derivatele parţiale ale funcţiei S în raport cu fiecare dintre cei doi parametri ai ecuaţiei de regresie (21.7), punând apoi condiţia ca derivatele parţiale să se anuleze şi efectuând calculele, rezultă sistemul:

$$\begin{cases} n\, lg\, a + b\sum_{i=1}^{n} lg\, x_i = \sum_{i=1}^{n} lg\, y_i \\ lg\, a\sum_{i=1}^{n} lg\, x_i + b\sum_{i=1}^{n} (lg\, x_i)^2 = \sum_{i=1}^{n} lg\, x_i\, lg\, y_i \end{cases} \tag{21.8}$$

Sistemul de ecuaţii normale (21.8) este aplicabil în primul caz de prezentare a informaţiilor iniţiale cu privire la cele două variabile şi se poate rezolva mai uşor cu ajutorul metodei lui *Cramer*. Formulele uzuale de calcul ale celor doi parametri ai ecuaţiei de regresie (21.7) vor lua următoarele expresii:

Scanned with OKEN Scanner

$$\lg a = \frac{\sum_{i=1}^{n}(\lg x_i)^2 \sum_{i=1}^{n} \lg y_i - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} \lg x_i \lg y_i}{n\sum_{i=1}^{n}(\lg x_i)^2 - (\sum_{i=1}^{n} \lg x_i)^2} \qquad (21.9)$$

şi

$$b = \frac{n\sum_{i=1}^{n} \lg x_i \lg y_i - \sum_{i=1}^{n} \lg x_i \sum_{i=1}^{n} \lg y_i}{n\sum_{i=1}^{n}(\lg x_i)^2 - (\sum_{i=1}^{n} \lg x_i)^2} \qquad (21.10)$$

21.3. Funcţiile putere cu doi factori

În cazul a două variabile factoriale X_1 şi X_2, funcţia putere (21.1) capătă forma cunoscută a funcţiilor de producţie *Cobb-Douglas* (21.4):

$$Y = aX_1^{b_1} X_2^{b_2} + \varepsilon$$

Parametrul ε reprezintă o variabilă aleatoare cu $m = 0$ şi $\sigma = 1$ şi care poate fi neglijată în aplicaţiile practice.

Această relaţie arată că o variaţie cu 1% a factorului de producţie X_1 determină o variaţie cu $b_1\%$ a volumului producţiei Y, atunci când factorul de producţie X_2 rămâne neschimbat, iar o variaţie cu 1% a factorului X_2 determină o variaţie egală cu $b_2\%$ a volumului producţiei, atunci când X_1 rămâne constant. Variaţia cu 1% a fiecăruia din cei doi factori de producţie X_1 şi X_2 determină o variaţie egală cu $(b_1 + b_2)\%$ a volumului producţiei.

Economiştii americani *Cobb* şi *Douglas* au aplicat funcţia de producţie cu două variabile la industria prelucrătoare a *SUA* ,bazându-se pe informaţiile statistice cu privire la producţia Y, capital X_1 şi forţa de muncă X_2 din perioada 1899-1922 şi au obţinut următoarea expresie:

Scanned with OKEN Scanner

$$Y = 1{,}01 \cdot X_1^{0{,}27} \cdot X_2^{0{,}73}$$

Ei au constatat, pentru prima dată, existenţa unei relaţii omogene liniare între volumul producţiei şi cei doi factori ai ei.

În modelul iniţial , parametrii b_1 şi b_2 constituie, deci, elasticităţi ale variabilelor factoriale X_1 şi X_2 şi arată cu cât se modifică variabila dependentă Y atunci când variabilele independente variază cu o unitate. Această interpretare devine evidentă calculând elasticităţile:

$$\frac{\partial Y}{\partial X_1} = b_1 a X_1^{b_1 - 1} X_2^{b_2}$$

iar

$$\frac{\partial Y}{\partial X_1} \cdot \frac{X_1}{Y} = \frac{b_1 a X_1^{b_1} X_2^{b_2}}{a X_1^{b_1} X_2^{b_2}} = b_1 \qquad (21.11)$$

respectiv

$$\frac{\partial Y}{\partial X_2} = b_2 a X_1^{b_1} X_2^{b_2 - 1}$$

$$\frac{\partial Y}{\partial X_2} \cdot \frac{X_2}{Y} = \frac{b_2 a X_1^{b_1} X_2^{b_2}}{a X_1^{b_1} X_2^{b_2}} = b_2 \qquad (21.12)$$

Cele două elasticităţi sunt complementare:

$$b_1 + b_2 = 1$$

iar

$$b_1 > 0 \quad \text{şi} \quad b_2 > 0 \qquad (21.13)$$

Aceasta înseamnă că funcţia putere de două variabile este o funcţie omogenă de gradul întâi.

Raportând ambii membri ai funcţiei de producţie (21.4) la numărul de lucrători sau la timpul consumat X_1, se va obţine o funcţie de productivitate a muncii:

$$\frac{Y}{X_1} = \frac{a X_1^{b_1} X_2^{b_2}}{X_1} = a \left(\frac{X_2}{X_1} \right)^{b_2} \qquad (21.14)$$

Scanned with OKEN Scanner

21.3.1. Reprezentarea grafică a funcţiilor putere cu doi factori

În cazul unei funcţii de producţie discrete, cu două variabile factoriale, graficul va lua forma unei diagrame prin paralelipipede dreptunghiulare (fig. 21.2).

Pentru o anumită cantitate de resurse consumate se obţine o anumită producţie, care este proporţională cu înălţimea fiecărui paralelipiped.

O funcţie de producţie *Cobb - Douglas* cu doi factori se poate reprezenta grafic şi sub forma unei suprafeţe poliedrale în cazul variabilelor continue şi poartă denumirea de *suprafaţă de producţie* (fig. 21.3). Un punct oarecare de pe această suprafaţă reprezintă valoarea producţiei care corespunde unui anumit volum de resurse utilizate. De exemplu, punctului I de pe suprafaţa de producţie îi corespunde un volum al producţiei dat de segmentul II', iar volumul resurselor utilizate este proporţional cu segmentele DI' şi $I'G$.

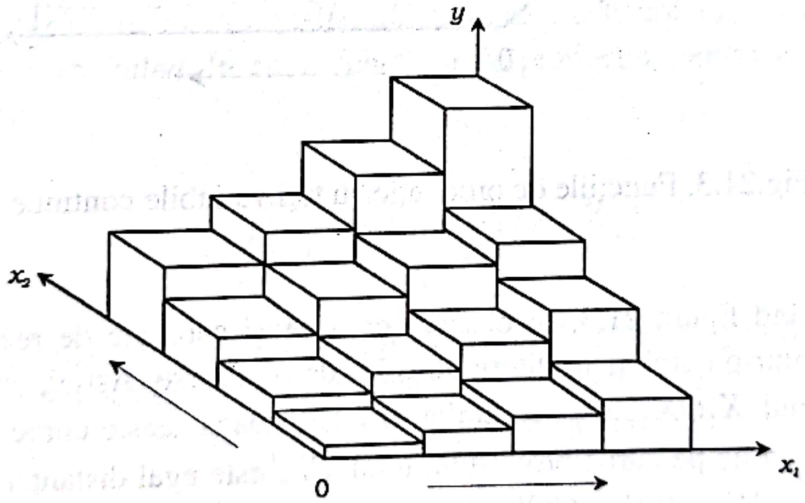


Fig.21.2. Funcţiile de producţie cu trei variabile discrete

Curba produsului total GIF înregistrează o creştere rapidă, corespunzător utilizării unor cantităţi reduse de fonduri de producţie până atinge un punct maxim, în care are loc o diminuare a produsului marginal al factorului - fonduri de producţie $\widetilde{P}_{x_1}$ corespunzătoare unui volum dat al forţei de muncă IG, după care $\widetilde{P}_K$ creşte în continuare cu o rată descrescândă.

În mod analog se interpretează şi curba produsului DIE.

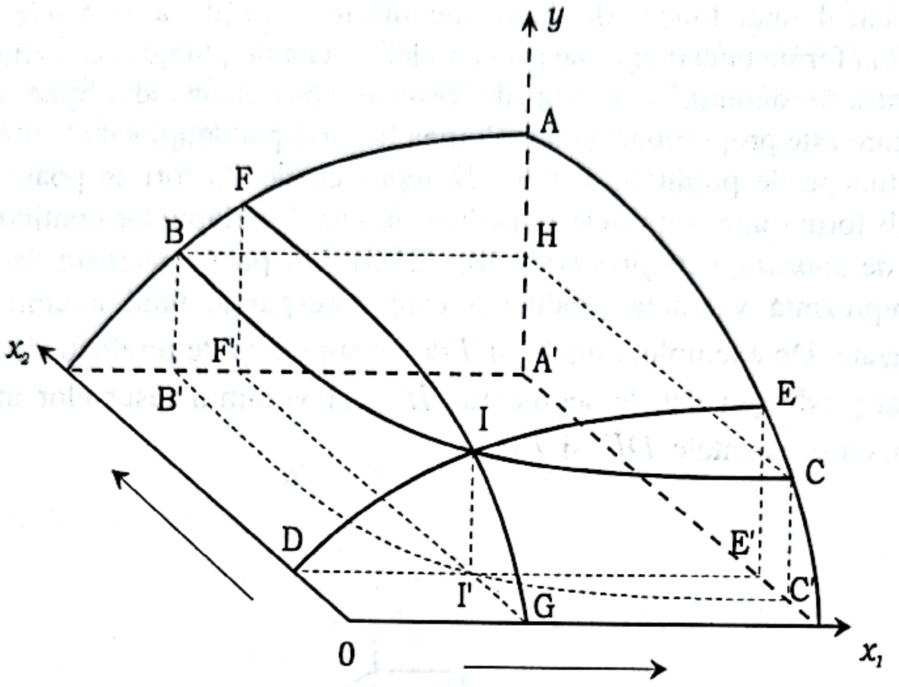


Fig.21.3. Funcţiile de producţie cu trei variabile continue

Analizând figura 21.3, se observă că aceeaşi cantitate de rezultate II' se poate obţine printr-o combinare diferită a celor două resurse. Astfel, curba BIC este paralelă cu planul X_1OX_2, deci cu curba $B'I'C'$. Dacă aceste curbe sunt paralele, atunci orice punct de pe curba produsului total BIC este egal distanţat faţă de curba $B'I'C'$, adică $II' = BB' = CC'$. Deci, fiecărui punct de pe curba BIC îi corespunde un punct de pe curba $B'I'C'$ şi acesta, la rândul lui, este determinat de o anumită abcisă şi o anumită ordonată în planul resurselor utilizate X_1OX_2.

Curbele de rezultate de tipul BIC poartă denumirea de *isoquante de producţie (production isoquants)*. Isoquantele de producţie arată combinaţiile posibile ale volumului celor două resurse pentru a produce o anumită cantitate de rezultate.

Dacă numărul variabilelor corelate este mai mare decât trei, expresia geometrică a funcţiei de producţie este o *hipersuprafaţă*, numită *colina producţiei* - obţinută într-un spaţiu multidimensional.

Scanned with OKEN Scanner

21.3.2. Proprietăţile funcţiilor de producţie

Ipoteze de lucru:

1. *Continuitatea.* În cazul variabilelor continue, funcţiile de producţie sunt nule, când cel puţin unul dintre factorii de producţie este nul:

$$f(X_1,0) = f(0,X_2) = f(0,0) = 0 \tag{21.15}$$

Relaţia (21.15) arată că ambii factori de producţie sunt indispensabili, pentru ca un proces de producţie să poată avea loc.

2. *Aditivitatea.* Funcţiile de producţie sunt aditive:

$$f(X_1' + X_1'', X_2' + X_2'') \geq f(X_1', X_2') + f(X_1'', X_2'') \tag{21.16}$$

Această ipoteză sugerează faptul că orice combinaţie a factorilor de producţie din cele două grupe considerate în (21.16) va duce la obţinerea unor rezultate cel puţin egale faţă de situaţia în care s-ar utiliza separat cele două grupe de factori.

3. *Divizibilitatea.* În cazul funcţiilor de producţie, pentru orice k pozitiv se verifică relaţia:

$$f\left(\frac{X_1}{k}, \frac{X_2}{k}\right) \geq \frac{1}{k} f(X_1, X_2) \tag{21.17}$$

Relaţia (21.17) semnifică faptul că procesul de producţie poate să aibă loc şi la o scară mai redusă, însă proporţia reducerii factorilor de producţie va fi mai mare decât proporţia reducerii rezultatului activităţii desfăşurate.

Proprietăţi

1. Dacă cele trei ipoteze de lucru se verifică, atunci rezultă că o majorare a factorilor de producţie de un anumit număr de ori λ va determina o creştere a producţiei de acelaşi număr de ori:

$$f(\lambda X_1, \lambda X_2) = \lambda f(X_1, X_2) \tag{21.18}$$

Într-adevăr,

$$Y = f(X_1, X_2) = a X_1^{b_1} X_2^{1-b_1} \quad \text{unde } a > 0$$

Scanned with OKEN Scanner

$$f(\lambda X_1, \lambda X_2) = a(\lambda X_1)^{b_1}(\lambda X_2)^{1-b_1} = a\lambda X_1^{b_1} X_2^{1-b_1} = \lambda f(X_1, X_2) = \lambda Y$$

Relaţia (21.18) se verifică în realitate, atunci când ceilalţi factori care acţionează asupra lui Y, în afară de factorii de producţie specificaţi X_1 şi X_2, rămân constanţi (în special cei tehnici).

Această proprietate a funcţiilor de producţie poartă denumirea de *proprietatea randamentului constant al factorilor*.

În cazul în care nu se verifică ipotezele din (21.16) şi (21.17), atunci, în locul randamentului constant, se va obţine un *randament crescător* sau *descrescător* dat de relaţia:

$$f(\lambda X_1, \lambda X_2) = \lambda^n f(X_1, X_2) \qquad (21.19)$$

Dacă: $n = 1$, se obţine un *randament constant*;

$n > 1$, se obţine un *randament crescător*;

$n < 1$, se obţine un *randament descrescător*.

Randamentul descrescător se poate obţine în realitate în cazurile în care producţia mai depinde, în afară de factorii specificaţi X_1 şi X_2 şi de unii factori rari, resurse limitate cum ar fi: pământul în agricultură, rezervele minerale în cazul industriei extractive etc.

Ţinând seama de teorema lui *Euler* asupra funcţiilor omogene, rezultă:

$$X_1 f'_{X_1}(X_1, X_2) + X_2 f'_{X_2}(X_1, X_2) = f(X_1, X_2) \qquad (21.20)$$

Împărţind ambii membri ai funcţiei de producţie printr-un factor de producţie , se obţine o funcţie de o singură variabilă:

$$\frac{Y}{X_1} = f\left(1, \frac{X_2}{X_1}\right) \qquad (21.21)$$

Substituind în relaţia (21.21) pe $\frac{X_2}{X_1} = Z$, rezultă:

$$\frac{Y}{X_1} = f(1, Z) \quad sau \quad Y = X_1 \varphi(Z) \qquad (21.22)$$

Scanned with OKEN Scanner

2. Funcţia de producţie continuă este crescătoare în raport cu X_1 şi X_2. Conform proprietăţii de aditivitate, rezultă:

$$f(X_1 + \Delta X_1, X_2) = f(X_1, X_2) + f(\Delta X_1, 0)$$

ceea ce înseamnă că funcţia creşte odată cu X_1, deoarece:

$$f(\Delta X_1, 0) = 0, \ iar \ f'_{X_1} \geq 0 \ şi \ f'_{X_2} \geq 0$$

Aceste derivate reprezintă *productivităţile marginale* ale factorilor de producţie, adică tocmai creşterea producţiei pe o unitate suplimentară dintr-un factor X_1 şi X_2.

În mod logic, productivităţile vor fi în descreştere în raport cu X_1 şi X_2, ceea ce înseamnă că o unitate suplimentară de capital sau de muncă este mult mai productivă când potenţialul existent este redus, decât atunci când potenţialul este ridicat. Practic, într-o întreprindere modernă cu procesul de producţie mecanizat şi automatizat într-un înalt grad, investiţiile suplimentare vor fi, fără îndoială, mai puţin eficiente decât într-un atelier cu muncă manuală.

21.3.3. Produs mediu şi produs marginal

Cu ajutorul funcţiilor de producţie, pot fi definite două noţiuni: produsul mediu şi produsul marginal.

În cazul funcţiilor de doi factori, *produsul mediu* $\overline{P}$ reprezintă cantitatea medie de produs obţinută pe o unitate de factor de producţie, când celălalt factor de producţie rămâne constant.

Dacă în funcţia de producţie $Y = f(X_1, X_2)$ considerăm valoarea lui X_2 dată, atunci:

$$Y = f(X_1/X_2) \ şi \ \overline{P} = \frac{Y}{X_1} = \frac{f(X_1/X_2)}{X_1} \qquad (21.23)$$

Produsul marginal $\widetilde{P}$ al unui factor de producţie reprezintă cantitatea suplimentară de produse obţinută prin creşterea cu o unitate a unui factor de producţie

când celălalt factor rămâne neschimbat. Produsul marginal se poate determina prin diferenţiere astfel:

$$\tilde{P}_{X_1} = \frac{\partial Y}{\partial X_1} = \frac{\partial f(X_1/X_2)}{\partial X_1} \quad sau \quad \tilde{P}_{X_1} = b_1 a X_1^{b_1-1} X_2^{b_2} \tag{21.24}$$

şi

$$\tilde{P}_{X_2} = \frac{\partial Y}{\partial X_2} = \frac{\partial f(X_2/X_1)}{\partial X_2} \quad sau \quad \tilde{P}_{X_2} = b_2 a X_1^{b_1} X_2^{b_2-1} \tag{21.25}$$

Deci, produsul marginal al unui factor de producţie este proporţional cu elasticitatea factorului respectiv şi cu produsul mediu al acestuia. În cazul funcţiilor de producţie, produsele marginale $\tilde{P}_{X_1}$ şi $\tilde{P}_{X_2}$ sunt pozitive, ceea ce înseamnă că Y este o funcţie crescătoare de X_1 şi X_2.

În cazul în care variabilele empirice se prezintă sub formă de mărimi relative, atât produsul mediu, cât şi produsul marginal sunt descrise de către o curbă continuă, crescătoare, la început, până atinge un punct maxim şi apoi descrescătoare.

În punctul său maxim valoarea produsului mediu este egală cu valoarea produsului marginal.

Astfel, fie o funcţie $Z = f(x)$, dacă notăm cu Δx creşterea variabilei x şi cu $f(x+\Delta x)$ noua valoare a funcţiei, atunci produsul marginal $\tilde{P}$ va fi:

$$\tilde{P} = \frac{f(x+\Delta x) - f(x)}{\Delta x} \tag{21.26}$$

sau folosind calculul diferenţial:

$$\frac{\partial Z}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \tag{21.27}$$

Aceasta înseamnă că , la limită, produsul marginal este tocmai panta curbei Z.

Produsul mediu şi produsul marginal al factorilor de producţie depind numai de proporţia în care sunt combinaţi factorii şi, în particular, ei sunt independenţi de volumul absolut al factorilor utilizaţi.

Pentru demonstrare considerăm o funcţie de producţie cu doi factori:

Scanned with OKEN Scanner

$$Y = f(X_1, X_2) = aX_1^{b_1} X_2^{b_2}$$

Împărţind ambii membri ai acestei relaţii prin X_1, rezultă produsul mediu pe seama lui X_1:

$$\overline{P}_{X_1} = \frac{Y}{X_1} = aX_1^{b_1-1} X_2^{1-b_1} = a\left(\frac{X_2}{X_1}\right)^{1-b_1} \qquad (21.28)$$

Procedând în mod similar pentru factorul X_2, rezultă:

$$\overline{P}_{X_2} = \frac{Y}{X_2} = aX_1^{b_1} X_2^{-b_1} = a\left(\frac{X_1}{X_2}\right)^{b_1} \qquad (21.29)$$

Pentru produsul marginal se poate obţine o relaţie asemănătoare.

Considerând în funcţia de producţie (21.4) pe X_2 constant, iar dacă X_1 va creşte până la $X_1 + \Delta X_1$, atunci:

$$\Delta Y = a(X_1 + \Delta X_1)^{b_1} X_2^{1-b_1} - aX_1^{b_1} X_2^{1-b_1} = aX_2^{1-b_1}[(X_1 + \Delta X_1)^{b_1} - X_1^{b_1}]$$

dar

$$(X_1 + \Delta X_1)^{b_1} = X_1^{b_1} + b_1 X_1^{b_1-1}\Delta X_1 + \frac{b_1(b_1-1)X_1^{b_1-2}(\Delta X_1)^2}{2!} + \cdots$$

Dacă în dezvoltarea acestui binom vom neglija valorile mici ale lui ΔX_1, respectiv acei termeni care cuprind puteri mai mari decât 1 ale lui ΔX_1, rezultă:

$$(X_1 + \Delta X_1)^{b_1} = X_1^{b_1} + b_1 X_1^{b_1-1}\Delta X_1$$

Substituind această relaţie în ΔY, rezultă:

$$\Delta Y \leftarrow aX_2^{1-b_1}(X_1^{b_1} + b_1 X_1^{b_1-1}\Delta X_1 - X_1^{b_1}) = b_1 aX_2^{1-b_1} X_1^{b_1-1}\Delta X_1$$

Produsul marginal al factorilor X_1:

Scanned with OKEN Scanner

$$\tilde{P}_{x_1} = \frac{\Delta Y}{\Delta X_1} = b_1 a X_2^{1-b_1} X_1^{b_1-1} = b_1 a \left(\frac{X_2}{X_1}\right)^{1-b_1} \qquad (21.30)$$

Deci produsul marginal al factorului X_1 depinde numai de proporţia acestui factor, fiind independent de volumul absolut al factorului respectiv.

Procedând în mod analog, se va obţine expresia produsului marginal al factorului X_2:

$$\Delta Y = aX_1^{b_1}(X_2 + \Delta X_2)^{1-b_1} - aX_1^{b_1} X_2^{1-b_1} = aX_1^{b_1}[(X_2 + \Delta X_2)^{1-b_1} - X_2^{1-b_1}]$$

iar

$$(X_2 + \Delta X_2)^{1-b_1} = X_2^{1-b_1} + (1-b_1)X_2^{-b_1}\Delta X_2$$

şi

$$\Delta Y = aX_1^{b_1}[X_2^{1-b_1} + (1-b_1)X_2^{-b_1}\Delta X_2 - X_2^{1-b_1}] = (1-b_1)aX_1^{b_1}X_2^{-b_1}\Delta X_2$$

În aceste condiţii, produsul marginal al factorului X_2 va fi:

$$\tilde{P}_{x_2} = \frac{\Delta Y}{\Delta X_2} = (1-b_1)aX_1^{b_1}X_2^{-b_1} = (1-b_1)a\left(\frac{X_1}{X_2}\right)^{b_1} \qquad (21.31)$$

21.3.4. Estimarea parametrilor unei funcţii putere cu doi factori

În vederea calculelor statistice ale parametrilor funcţiei putere de două variabile factoriale (21.4), se procedează la liniarizarea acestui model prin logaritmare şi se obţine:

$$\lg Y = \lg a + b_1 \lg X_1 + b_2 \lg X_2 \qquad (21.32)$$

Parametrii ecuaţiei (21.32) se pot estima prin utilizarea metodei celor mai mici pătrate astfel:

Scanned with OKEN Scanner

$$S = \sum_{j=1}^{n} (lg\, y_j - lg\, y_{x_1,x_2})^2 = minim$$

sau

$$S = \sum_{j=1}^{n} (lg\, y_j - lg\, a - b_1\, lg\, x_1 - b_2\, lg\, x_2)^2 = minim$$

Determinând derivatele parţiale în raport cu parametrii a, b_1, b_2 şi efectuând calculele elementare, rezultă un sistem de trei ecuaţii normale:

$$\begin{cases} n\,lg\,a + b_1 \sum_{j=1}^{n} lg\, x_{1j} + b_2 \sum_{j=1}^{n} lg\, x_{2j} = \sum_{j=1}^{n} lg\, y_j \\ lg\,a \sum_{j=1}^{n} lg\, x_{1j} + b_1 \sum_{j=1}^{n} (lg\, x_{1j})^2 + b_2 \sum_{j=1}^{n} lg\, x_{1j}\, lg\, x_{2j} = \sum_{j=1}^{n} lg\, x_{1j}\, lg\, y_j \\ lg\,a \sum_{j=1}^{n} lg\, x_{2j} + b_1 \sum_{j=1}^{n} lg\, x_{1j}\, lg\, x_{2j} + b_2 \sum_{j=1}^{n} (lg\, x_{2j})^2 = \sum_{j=1}^{n} lg\, x_{2j}\, lg\, y_j \end{cases} \quad (21.33)$$

În rezolvarea acestui sistem trebuie să se ţină seama de modul de sistematizare a informaţiilor iniţiale cu privire la cele trei variabile corelate, iar dacă variabilele se prezintă sub forma unor serii dinamice, va trebui să se acorde importanţa cuvenită şi depistării şi eliminării fenomenului de multicoliniaritate.

Sistemul (21.33) se poate aplica în practică, în cazul în care variabilele corelate se prezintă sub forma unor şiruri de variante neponderate.

Dacă variabilele se prezintă sub forma unor distribuţii empirice, este necesar să se ţină seama în calcule de frecvenţele variantelor variabilelor. În acest caz, sistemul de ecuaţii normale (21.33) devine:

$$\begin{cases} lg\,a \sum_{j=1}^{n} n_j + b_1 \sum_{j=1}^{n} n_j\, lg\, x_{1j} + b_2 \sum_{j=1}^{n} n_j\, lg\, x_{2j} = \sum_{j=1}^{n} n_j\, lg\, y_j \\ lg\,a \sum_{j=1}^{n} n_j\, lg\, x_{1j} + b_1 \sum_{j=1}^{n} n_j (lg\, x_{1j})^2 + b_2 \sum_{j=1}^{n} n_j\, lg\, x_{1j}\, lg\, x_{2j} = \sum_{j=1}^{n} n_j\, lg\, x_{1j}\, lg\, y_j \\ lg\,a \sum_{j=1}^{n} n_j\, lg\, x_{2j} + b_1 \sum_{j=1}^{n} n_j\, lg\, x_{1j}\, lg\, x_{2j} + b_2 \sum_{j=1}^{n} n_j (lg\, x_{2j})^2 = \sum_{j=1}^{n} n_j\, lg\, x_{2j}\, lg\, y_j \end{cases} \quad (21.34)$$

Scanned with OKEN Scanner

Dacă variabilele modelului sunt grupate sub forma unui tabel cu dublă intrare, este necesar să se ia în calcul frecvenţele perechilor de valori corelate.

21.4. Funcţia putere cu trei factori

În anul 1936, *Edelberg* a introdus a treia variabilă independentă în funcţiile de producţie şi anume, durata procesului de producţie, astfel:

$$Y = aX_1^{b_1} X_2^{b_2} X_3^{b_3}, \quad unde \quad b_1 + b_2 + b_3 \le 1 \tag{21.35}$$

Prin logaritmare ecuaţia (21.35) devine:

$$lg\, Y_j = lg\, a + b_1\, lg\, X_1 + b_2\, lg\, X_2 + b_3\, lg\, X_3 \tag{21.36}$$

Aplicând metoda celor mai mici pătrate se obţine sistemul de ecuaţii normale:

$$\begin{cases} n\lg a + b_1\sum_{j=1}^{n}\lg x_{1j} + b_2\sum_{j=1}^{n}\lg x_{2j} + b_3\sum_{j=1}^{n}\lg x_{3j} = \sum_{j=1}^{n}\lg y_j \\ \lg a\sum_{j=1}^{n}\lg x_{1j} + b_1\sum_{ji=1}^{n}(\lg x_{1j})^2 + b_2\sum_{i=1}^{n}\lg x_{1j}\lg x_{2j} + b_3\sum_{j=1}^{n}\lg x_{1j}\lg x_{3j} = \sum_{j=1}^{n}\lg x_{1j}\lg y_j \\ \lg a\sum_{j=1}^{n}\lg x_{2j} + b_1\sum_{j=1}^{n}\lg x_{1j}\lg x_{2j} + b_2\sum_{i=1}^{n}(\lg x_2)^2 + \sum_{i=1}^{n}\lg x_{2j}\lg x_{3j} = \sum_{j=1}^{n}\lg x_{2j}\lg y_j \\ \lg a\sum_{j=1}^{n}\lg x_{3j} + b_1\sum_{j=1}^{n}\lg x_{1j}\lg x_{3j} + b_2\sum_{j=1}^{n}\lg x_{2j}\lg x_{3j} + b_3\sum_{j=1}^{n}(\lg x_{3j})^2 + = \sum_{j=1}^{n}\lg x_{3j}\lg y_j \end{cases} \tag{21.37}$$

Utilizarea în practică a modelului funcţiilor de producţie ridică o serie de probleme din punct de vedere statistic şi economic.

Principalul obstacol din punct de vedere statistic îl constituie fenomenul de multicoliniaritate între capital şi forţa de muncă, deoarece ambii au o tendinţă de creştere în timp. Mai mult, dacă seriile dinamice reprezentând variabilele independente sunt afectate şi de ciclicitate, valorile parametrilor de regresie b_i riscă să fie determinate mai mult de fluctuaţiile ciclice, decât de variaţia unui fenomen pe termen lung.

Pentru rezolvarea acestei probleme, literatura de specialitate propune două soluţii: fie înlocuirea variabilelor prezentate sub formă de serii dinamice, cu

Scanned with OKEN Scanner

observaţii instantanee provenind din diferite ţări sau diferite unităţi administrativ teritoriale ale aceleiaşi ţări, fie analiza teoretică a factorilor de producţie şi stabilirea pe această bază a raportului $\frac{b_2}{b_1} = K$ şi apoi să se ajusteze funcţia:

$$Y = aX_1^{b_1}X_2^{Kb_1} \quad sau \quad Y = a(X_1X_2^K)^{b_1}$$

Din punct de vedere economic principalele critici se referă la faptul că, în funcţiile de producţie iniţiale, se luau în calcul prea puţini factori de producţie şi, în plus, nu se ţinea seama de aspectele calitative ale acestora, cum ar fi calitatea utilizării factorilor de producţie. Astfel două întreprinderi cu acelaşi capital şi acelaşi număr de muncitori pot avea producţii foarte diferite. De asemenea, modelul iniţial este invariabil în timp şi nu ţine seama de progresul tehnic.

În vederea soluţionării acestor deficienţe ale funcţiilor de producţie, s-a propus ca variabila X_1 şi parametrii b_1 şi b_2 să fie funcţii de timp, astfel:

$$Y = A_0X_1^{a_0+a_1t}X_2^{c_0+c_1t}e^{rt}$$

unde r este o constantă care caracterizează ritmul de creştere.

Logaritmând această expresie, se obţine o ecuaţie comodă pentru calculele statistice:

$$lgY = lgA_0 + (a_0+a_1t)lgX_1 + (c_0+c_1t)lgX_2 + rt$$

Sau, mai simplu, presupunând că b_1 şi b_2 rămân constanţi şi că variază numai a se obţine:

$$Y = aX_1^{b_1}X_2^{b_2}e^{rt}$$

Aceste modificări ale funcţiei de producţie arată că, având un stoc de capital dat şi acelaşi număr de lucrători, eficacitatea utilizării factorilor are o tendinţă de creştere în timp, prin mai buna utilizare a tehnicii şi muncii (learning by doing), iar tehnicile noi, mai moderne, care le înlocuiesc pe cele vechi cu acelaşi număr de lucrători şi aceeaşi valoare a dotării asigură obţinerea unui plus de producţie.

Factorul e^{rt} reprezintă, într-un anume fel, în mod sintetic progresul tehnic.

Scanned with OKEN Scanner

VI. ANALIZA VARIATIEI FENOMENELOR IN TIMP

Scanned with OKEN Scanner

22. INDICATORII SERIILOR DINAMICE

23. TENDINŢA GENERALĂ DE DEZVOLTARE A FENOMENELOR ÎN TIMP

24. STUDIUL VARIAŢIILOR SEZONIERE, CICLICE ŞI A CELOR ALEATOARE

Scanned with OKEN Scanner

22. INDICATORII SERIILOR DINAMICE

Timpul şi spaţiul reprezintă cele două forme fundamentale de existenţă a materiei. Dezvoltarea oricărui fenomen din natură sau din societate are loc într-un anumit spaţiu care îşi imprimă asupra fenomenelor anumite trăsături specifice şi într-un anumit moment sau interval de timp, caracteristică care, de asemenea, imprimă fenomenelor anumite proprietăţi.

Termenii seriilor dinamice prezintă nivelul de dezvoltare al fenomenelor în anumite intervale de timp bine determinate, şi de regulă egale între ele, pentru a se putea face comparaţie între intervale sau nivelul atins în diferite momente de timp din dezvoltarea lor. Intervalele dintre momente pot fi egale sau neegale între ele.

Corespunzător principalelor feluri de mărimi utilizate în statistică, indicatorii seriilor dinamice se pot împărţi în următoarele *categorii*:

Indicatorii seriilor dinamice:

- *indicatori absoluţi:*
  - *nivelul absolut;*
  - *sporul absolut:*
    - *cu bază fixă;*
    - *cu bază mobilă;*
- *indicatori relativi:*
  - *indicele variaţiei totale:*
    - *cu bază fixă;*
    - *cu bază mobilă;*
  - *indicele sporului:*
    - *cu bază fixă;*
    - *cu bază mobilă;*
- *indicatori medii:*
  - *nivelul mediu;*
  - *sporul mediu absolut;*
  - *indicele mediu al variaţiei;*
  - *ritmul mediu al variaţiei.*

Scanned with OKEN Scanner

22.1. Indicatorii nivelului absolut şi cei ai sporului

Pentru prezentarea metodelor de calcul şi analiză statistică a indicatorilor nivelului şi sporului absolut al seriilor dinamice (de momente sau de intervale) considerăm seria dinamică din tabelul 22.1.

Tabelul 22.1. Indicatorii seriilor dinamice exprimaţi în mărimi absolute

Timpul t_i	Nivelurile y_i	Sporul absolut	
		Cu bază fixă $\Delta_{y_i/0}$	Cu baza mobilă $\Delta_{y_i/i-1}$
1	2	3	4
t_0	y_0	-	-
t_1	y_1	$\Delta_{y_1/0}$	$\Delta_{y_1/0}$
t_2	y_2	$\Delta_{y_2/0}$	$\Delta_{y_2/1}$
⋮	⋮	⋮	⋮
t_i	y_i	$\Delta_{y_i/0}$	$\Delta_{y_i/i-1}$
⋮	⋮	⋮	⋮
t_n	y_n	$\Delta_{y_n/0}$	$\Delta_{y_n/n-1}$
Total	$\sum_{i=0}^{n} y_i$	-	$\Delta_{y_n/0}$

Sporul sau excedentul absolut Δ_y măsoară în mărimi absolute variaţia unui fenomen într-un anumit interval de timp şi se calculează ca o diferenţă între nivelul atins de un fenomen la un moment dat şi nivelul aceluiaşi fenomen dintr-un alt moment anterior, considerat ca bază de comparaţie. Termenul ales ca bază de comparaţie trebuie să reprezinte un nivel normal din dezvoltarea fenomenului luat în calcul, supus într-o măsură cât mai mică influenţei factorilor perturbatori ai dezvoltării.

În funcţie de modul de alegere a bazei de comparaţie într-o serie dinamică se disting două categorii de spor absolut:

Scanned with OKEN Scanner

- un spor absolut cu bază fixă - notat cu $\Delta_{y_{i/0}}$;
- un spor absolut cu bază mobilă - notat cu $\Delta_{y_{i/i-1}}$.

Sporul absolut cu bază fixă se caracterizează prin faptul că nivelul considerat ca bază rămâne acelaşi pentru întreaga serie. Dacă considerăm ca nivel de bază primul termen al unei serii dinamice y_0, atunci sporul absolut cu bază fixă se poate calcula cu ajutorul formulei:

$$\Delta_{y_{i/0}} = y_i - y_0 \qquad i = 0,1,2,\ldots,n \tag{22.1}$$

În domeniul economic se consideră ca bază de comparaţie un nivel din dezvoltarea unui fenomen cu o semnificaţie aparte. De exemplu, pentru a caracteriza evoluţia producţiei unor produse industriale principale pe o anumită durată de timp, se compară nivelurile atinse în diferite etape ale perioadei cu nivelul aceluiaşi fenomen înregistrat la sfârşitul perioadei precedente. În toate cazurile alegerea bazei de comparaţie nu trebuie să se facă la întâmplare, ci numai în urma unei analize temeinice a evoluţiei fenomenelor în timp şi în funcţie de scopul (cunoaşterea ştiinţifică) urmărit în fiecare caz în parte. Precizăm, de asemenea, că baza de comparaţie nu trebuie să fie prea mult depărtată în timp de nivelul care se compară deoarece dezvoltarea în timp înseamnă nu numai o creştere (sau scădere) cantitativă a aceleiaşi calităţi, ci şi modificări calitative, de esenţă, uneori chiar foarte importante, care modifică însăşi natura fenomenelor şi le fac să nu mai fie întru totul comparabile.

Sporul absolut cu bază mobilă se caracterizează prin faptul că baza de comparaţie se schimbă în mod continuu, de la un termen la altul, considerându-se de fiecare dată drept bază nivelul perioadei precedente. Astfel, se caracterizează variaţia fenomenelor de la o zi la alta, de la o săptămână la alta, de la o lună la alta, de la un an la altul etc.

Formula de calcul al sporului absolut cu bază mobilă sau în lanţ, cum i se mai spune, este:

$$\Delta_{y_{i/i-1}} = y_i - y_{i-1} \qquad i = \overline{1,n} \tag{22.2}$$

Scanned with OKEN Scanner

Între sporurile cu bază fixă şi sporurile cu bază mobilă există anumite relaţii bine determinate, care permit trecerea de la o categorie de sporuri la cealaltă fără a se recurge la seria dinamică pe baza căreia au fost calculate aceste sporuri:

1. *Suma sporurilor cu bază mobilă, dintr-o anumită perioadă de timp, este egală cu sporul cu bază fixă al întregii perioade*, adică:

$$\sum_{i=1}^{n} \Delta_{y_{i/i-1}} = \Delta_{y_{n/0}} \tag{22.3}$$

$$\sum_{i=1}^{n} \Delta_{y_{i/i-1}} = (y_1 - y_0) + (y_2 - y_1) + \cdots + (y_i - y_{i-1}) + \cdots + (y_n - y_{n-1}) = y_n - y_0 = \Delta_{y_{n/0}}$$

2. *Diferenţa dintre sporurile absolute cu bază fixă consecutive este egală cu sporul cu bază mobilă corespunzător*:

$$\Delta_{y_{i/0}} - \Delta_{y_{i-1/0}} = \Delta_{y_{i/i-1}} \qquad i = \overline{1,n} \tag{22.4}$$

deoarece

$$\Delta_{y_{i/0}} - \Delta_{y_{i-1/0}} = (y_i - y_0) - (y_{i-1} - y_0) = y_i - y_{i-1} = \Delta_{y_{i/i-1}}$$

Pentru ilustrarea calculului şi analizei statistice a sporului absolut al fenomenelor, să considerăm seria dinamică din tabelul 22.2., reprezentând creşterea producţiei de energie electrică în România în anii 1950, 1960, 1980, 1990, 1994.

Un alt indicator al variaţiei fenomenelor, exprimat în aceleaşi unităţi de măsură ca şi fenomenul însuşi, este *sporul mediu* $\overline{\Delta}_y$. Sporul mediu arată în mărimi absolute cât de mare ar fi fost variaţia (creşterea sau descreşterea) unui fenomen în perioade mai mici de timp dar egale şi cuprinse în cadrul perioadei mai mari luate în calcul.

Sporul mediu se poate calcula ca o medie aritmetică simplă a sporurilor cu bază mobilă după formula:

Scanned with OKEN Scanner

$$\overline{\Delta}_y = \frac{\sum_{i=1}^{n} \Delta_{y_{i/i-1}}}{n} \qquad (22.5)$$

Tabelul 22.2. Producţia de energie electrică şi sporul absolut în anii 1950, 1960, 1970, 1980, 1990 şi 1995

Anii	Milioane kwh	Sporul absolut					
		Cu baza fixă $\Delta_{y\%}$					Cu baza mobilă $\Delta_{y_{i/i-1}}$
	y_i	y_{1950}	y_{1960}	y_{1970}	y_{1980}	y_{1990}	
1950	2113	0	-5537	-32975	-65373	-62196	-
1960	7650	5537	0	-27438	-59836	-56659	5537
1970	35088	32975	27438	0	-32398	-29221	27438
1980	67486	65373	59836	32398	0	3177	32398
1989	75851	73746	68209	40771	8365	-11542	8365
1990	64309	62196	56659	29221	3177	0	-11542
1991	56912	54799	49262	21284	10574	- 7397	- 8000
1992	54195	52082	46545	19107	13291	-10114	- 2717
1993	55476	53363	47826	20388	12010	- 8833	1281
1994	55136	53023	47486	20048	12350	- 9173	- 340
1995	59267	57154	51617	24179	- 8219	- 5042	4131

Sursa: Preluat după *Anuarele Statistice ale României*

În baza relaţiei (22.3) formula sporului mediu devine:

$$\overline{\Delta}_y = \frac{\Delta_{y_{n/0}}}{n} \qquad (22.6)$$

Relaţia (22.6) arată că sporul mediu se poate calcula şi ca un raport între sporul cu bază fixă al întregii perioade şi numărul termenilor seriei minus 1. În formula (22.6), ca şi în cele anterioare, s-a considerat că seria pentru care se calculează indicatorii sporului, este formată din $n+1$ termeni.

Scanned with OKEN Scanner

Dacă în relaţia (22.6) se substituie sporul cu bază fixă al întregii perioade cu expresia lui din (22.1), rezultă formula uzuală de calcul al sporului mediu:

$$\overline{\Delta}_y = \frac{y_n - y_0}{n} \qquad (22.7)$$

Sporul mediul anual de creştere a producţiei de energie electrică pentru perioada 1950-1995 a avut valorile următoare:

$$\overline{\Delta}_{y_{1950-1995}} = \frac{59267 - 2113}{50} = 1143{,}1 \text{ milioane kwh/an}$$

Aceasta înseamnă că în perioada 1950-1995 producţia de energie electrică a crescut în medie, în fiecare an, cu 1143,1 milioane kwh .

Pentru diferite subperioade s-au obţinut sporurile medii anuale, diferite astfel:

$$\overline{\Delta}_{y_{1950-1960}} = \frac{5537}{10} = 553{,}7 \textit{ milioane kwh}$$

$$\overline{\Delta}_{y_{1960-1970}} = \frac{27438}{10} = 2743{,}8 \textit{ milioane kwh}$$

$$\overline{\Delta}_{y_{1970-1980}} = \frac{32398}{10} = 3239{,}8 \textit{ milioane kwh}$$

$$\overline{\Delta}_{y_{1980-1990}} = \frac{-3177}{10} = -317{,}7 \textit{ milioane kwh}$$

$$\overline{\Delta}_{y_{1990-1995}} \frac{-5042}{5} = -100{,}8 \textit{ milioane kwh}$$

Rezultă că în *România* în perioada 1950-1995 a avut loc o creştere continuă a producţiei medii anuale de energie electrică până în anul 1989, după

Scanned with OKEN Scanner

care aceasta s-a diminuat în următorii trei ani pentru ca apoi, în perioada 1993-1995 să asistăm la o tendinţă de creştere a acesteia.

Din formula (22.7) se observă că sporul mediu depinde numai de mărimea termenilor extremi ai unei serii şi de numărul lor, făcând abstracţie de toţi termenii intermediari. Aceasta înseamnă că metoda clasică de determinare a sporului mediu exagerează artificial importanţa primului şi a ultimului termen ai unei serii şi-i neglijează pe toţi ceilalţi. Aceasta reprezintă principala deficienţă a metodei obişnuite de calcul al sporului mediu - deficienţă care-i limitează sfera de aplicabilitate practică numai la seriile dinamice în care termenii reprezintă niveluri normale în dezvoltarea fenomenului studiat. În caz contrar toate rezultatele, obţinute prin prelucrarea statistică ulterioară, se vor abate faţă de realitate aproximativ în aceeaşi măsură în care termenii y_0 şi y_n, sau ambii termeni se abat de la dezvoltarea normală a fenomenului.

De aceea, ori de câte ori se constată că unul sau ambii termeni extremi se abat de la dezvoltarea normală a fenomenului cercetat, pentru a se înlătura deficienţa arătată, se recomandă să se calculeze sporul mediu cu ajutorul metodei celor mai mici pătrate, care ţine seama de toate nivelurile individuale ale seriei.

În acest scop să presupunem că dezvoltarea unui fenomen are loc în mod liniar, după funcţia:

- în momentul t :

$$y = a + bt$$

- în momentul $(t + \Delta t)$:

$$y + \Delta_y = a + b(t + \Delta_t)$$

$$y + \Delta_y = a + bt + b\Delta_t$$

Sporul mediu poate fi estimat ca un raport între creşterea funcţiei şi creşterea argumentului.

Expresia geometrică a parametrului b se prezintă sugestiv în figura 22.1.

Scanned with OKEN Scanner

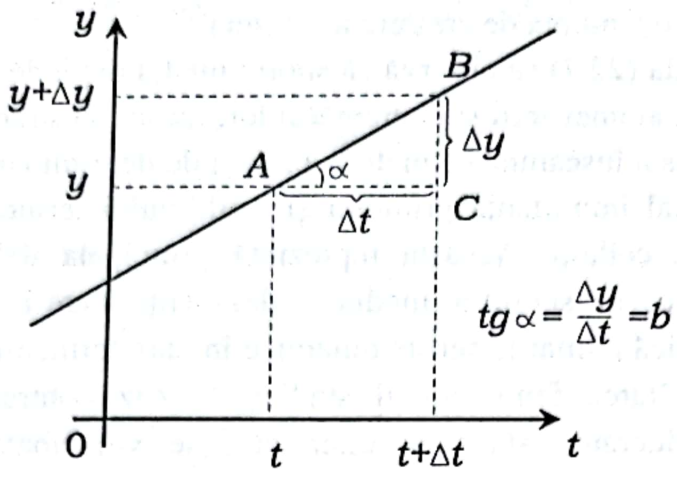


Figura 22.1. Determinarea sporului mediu

Scăzând din valoarea funcţiei în momentul $t+\Delta t$ valoarea ei în momentul t, rezultă:

$$\Delta_y = b\Delta_t \quad \text{iar} \quad b = \frac{\Delta_y}{\Delta_t} \tag{22.8}$$

Din relaţia (22.8) se observă că parametrul b din ecuaţia unei drepte reprezintă viteza de creştere a funcţiei liniare sau sporul mediu.

Sporul mediu este un indicator sintetic al variaţiei fenomenelor cu importante implicaţii în analiza şi prognoza fenomenelor pe baza seriilor dinamice.

22.2. Indicele variaţiei totale a fenomenelor

Un alt indicator important al variaţiei fenomenelor în timp îl reprezintă viteza de variaţie, exprimată în mărimi relative, numit indice de variaţie sau ritm al variaţiei totale. Ca orice mărime relativă şi indicele variaţiei în timp se obţine prin compararea nivelului unui fenomen, atins într-un anumit moment, cu un alt nivel

Scanned with OKEN Scanner

anterior al aceluiaşi fenomen, considerat ca bază de comparaţie. În funcţie de modul de alegere a bazei de comparaţie indicele variaţiei este de două feluri:

- indicele variaţiei fenomenelor cu bază fixă;
- indicele variaţiei fenomenelor cu bază mobilă.

Indicele variaţiei cu bază fixă, arată de câte ori a crescut (sau a scăzut) nivelul unui anumit fenomen în decursul unei anumite perioade de timp şi se calculează după formula:

$$I_{y_{i/0}} = \frac{y_i}{y_0} \qquad i = 0,1,2,...,n \qquad (22.9)$$

Indicele variaţiei cu bază mobilă se determină cu ajutorul formulei:

$$I_{y_{i/i-1}} = \frac{y_i}{y_{i-1}} \qquad i = 0,1,2,...,n \qquad (22.10)$$

şi arată de câte ori a crescut sau a scăzut nivelul unui anumit fenomen într-un anumit moment sau interval de timp faţă de un moment sau perioadă anterioară.

Consideraţiile privitoare la alegerea bazei de comparaţie în cazul sporului absolut sunt aplicabile şi indicelui variaţiei.

Relaţiile dintre indicii variaţiei cu bază fixă şi cei cu bază mobilă se pot enunţa astfel:

1. *Produsul indicilor variaţiei cu bază mobilă este egal cu indicele variaţiei cu bază fixă al întregii perioade:*

$$\prod_{i=1}^{n} I_{y_{i/i-1}} = I_{y_{n/0}} \qquad (22.11)$$

Într-adevăr,

$$\prod_{i=1}^{n} I_{y_{i/i-1}} = \frac{y_1}{y_0} \cdot \frac{y_2}{y_1} \cdots \frac{y_i}{y_{i-1}} \cdots \frac{y_n}{y_{n-1}} = \frac{y_n}{y_0} = I_{y_{n/0}}$$

2. *Raportul între doi indici ai variaţiei cu bază fixă este egal cu indicele cu bază mobilă corespunzător:*

$$I_{y_{i/0}} : I_{y_{i-1/0}} = I_{y_{i/i-1}} \qquad (22.12)$$

Scanned with OKEN Scanner

22.3. Indicele creşterii fenomenelor

Indicele creşterii fenomenelor, numit şi ritmul creşterii, exprimă mărimea creşterii (sau scăderii) lor în decursul unei anumite perioade de timp faţă de nivelul considerat ca bază.

După modul de alegere a bazei de comparaţie se disting două feluri de indici ai creşterii, numiţi şi indici ai sporului:

- indicii creşterii cu bază fixă $r_{y_{i/0}}$;
- indicii creşterii cu bază mobilă $r_{y_{i/i-1}}$.

Indicii creşterii cu bază fixă se calculează prin raportarea sporului cu bază fixă la nivelul fenomenului considerat ca bază de comparaţie, după formula:

$$r_{y_{i/0}} = \frac{\Delta_{y_{i/0}}}{y_0} \qquad (22.13)$$

sau

$$r_{y_{i/0}} = \frac{y_i - y_0}{y_0}.$$

Pentru exemplificare să considerăm din nou seria din tabelul 22.2.

Tabelul 22.3. Producţia de energie electrică şi sporul absolut în anii 1950, 1960, 1970, 1980, 1990, 1990-1995

Anii	Mil. kwh	Indicii variaţiei cu bază fixă					Indicii variaţiei cu bază mobilă
	y_i 60	y_{1950}	y_{1960}	y_{1970}	y_{1980}	y_{1990}	
1950	2113	**1,00**	0,28	0,06	0,03	0,03	1,87
1960	7650	3,62	**1,00**	0,22	0,11	0,10	3,62
1970	35088	16,61	4,59	**1,00**	0,52	0,46	4,59
1980	67486	31,94	8,82	1,92	**1,00**	0,89	1,92
1989	75851	35,90	9,92	2,16	1,12	**1,00**	1,12
1990	64309	30,43	8,41	1,83	0,95	0,85	0,85
1991	56912	26,93	7,44	1,62	0,84	0,75	0,88
1992	54195	25,65	7,08	1,541	0,80	0,71	0,95
1993	55476	26,25	7,25	,58	0,82	0,73	1,02
1994	55136	26,09	7,21	1,57	0,82	0,73	0,99
1995	59267	28,05	7,75	1,69	0,88	0,92	1,07

Scanned with OKEN Scanner

Indicii creşterii cu bază mobilă se pot calcula prin raportarea sporului cu bază mobilă la nivelul fenomenului din perioada de bază, astfel:

$$r_{y_{i/i-1}} = \frac{\Delta_{y_{i/i-1}}}{y_{i-1}} \quad \text{sau} \quad r_{y_{i/i-1}} = \frac{y_i - y_{i-1}}{y_{i-1}} \qquad (22.14)$$

Între indicii variaţiei totale şi indicii creşterii există o relaţie bine determinată şi anume: *indicele creşterii (sporului) este egal cu indicele variaţiei totale minus 1 sau 100 în funcţie de modul de exprimare a acestor indici sub formă de coeficient sau în procente.*

Astfel:

$$r_{y_{i/0}} = I_{y_{i/0}} - 1 \qquad (22.15)$$

şi

$$r_{y_{i/i-1}} = I_{y_{i/i-1}} - 1 \qquad (22.16)$$

Valoarea absolută a unui procent de creştere cu bază fixă $V_{y_{i/0}}$ se obţine prin raportarea sporului absolut cu bază fixă $\Delta_{y_{i/0}}$ la ritmul sporului $r_{y_{i/0}}$ cu bază fixă al aceleiaşi perioade:

$$V_{y_{i/0}} = \frac{\Delta_{y_{i/0}}}{r_{y_{i/0}}(\%)} = \frac{y_i - y_0}{\frac{y_i - y_0}{y_0} \cdot 100} = \frac{y_0}{100} \qquad i = \overline{1,n}$$

Aceasta înseamnă că valoarea absolută a unui procent de creştere cu bază fixă este aceeaşi pentru fiecare perioadă i .

Valoarea absolută a unui procent de creştere cu bază mobilă $V_{y_{i/i-1}}$ se determină prin raportarea sporului absolut cu bază mobilă $\Delta_{y_{i/i-1}}$ al perioadei la ritmul sporului (creşterii) cu bază mobilă $r_{y_{i/i-1}}$ al aceleiaşi perioade:

$$V_{y_{i/i-1}} = \frac{\Delta_{y_{i/i-1}}}{r_{y_{i/i-1}}(\%)} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100} = \frac{y_{i-1}}{100} \qquad i = \overline{1,n}$$

Scanned with OKEN Scanner

Din această relaţie rezultă că valoarea absolută a unui procent de creştere cu bază mobilă este diferită de la o perioadă la alta, fiind egală de fiecare dată cu a suta parte din nivelul fenomenului considerat ca bază y_{i-1}.

22.3.1. Indicele mediu de creştere

Indicele mediu de creştere exprimă viteza medie relativă de creştere a unui fenomen în decursul unei anumite perioade de timp. Astfel, se poate calcula indicele mediu anual de creştere a unui fenomen în decursul unei perioade de 5 ani, 10 ani etc. Indicele mediu anual măsoară creşterea medie anuală a unui fenomen în ipoteza că, în decursul întregii perioade, fenomenul s-ar fi dezvoltat cu aceeaşi creştere relativă.

În teoria şi practica statistică se cunosc mai multe metode de determinare a indicelui mediu de creştere a fenomenelor, dintre care, mai des utilizate sunt: *metoda mediei geometrice, metoda mediei parabolice şi metoda trendului exponenţial.*

22.3.1.1. Metoda mediei geometrice

În ipoteza dezvoltării unui fenomen cu indici de variaţie cu bază mobilă constanţi, după modelul unei progresii geometrice, se poate determina *indicele mediu de creştere* ca o raţie a progresiei geometrice. În această ipoteză termenii seriei dinamice se pot scrie astfel:

$$\begin{aligned} y_0 &= y_0 \bar{I}_y^0 \\ y_1 &= y_0 \bar{I}_y^1 \\ y_2 &= y_0 \bar{I}_y^2 \\ &\cdots\cdots\cdots \\ y_i &= y_0 \bar{I}_y^i \\ &\cdots\cdots\cdots \\ y_n &= y_0 \bar{I}_y^n \end{aligned} \qquad (22.17)$$

Scanned with OKEN Scanner

Indicele mediu de creştere $\bar{I}_y$ se poate calcula pe baza relaţiei (22.17) astfel:

$$\bar{I}_y^n = \frac{y_n}{y_0}$$

iar

$$\bar{I}_y = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[n]{I_{n/0}} \qquad (22.18)$$

Relaţia (22.18) arată că indicele mediu de creştere a unui anumit fenomen se poate determina prin extragerea rădăcinii de un ordin egal cu numărul termenilor seriei minus *1* din indicele variaţiei cu bază fixă al întregii perioade.

Ţinând seama de relaţia (22.11) formula indicelui mediu de creştere devine:

$$\bar{I}_y = \sqrt[n]{\prod_{i=1}^{n} I_{i/i-1}} \qquad (22.19)$$

Din relaţia (22.19) rezultă că indicele mediu de creştere a fenomenelor se poate determina ca o medie geometrică a indicilor de variaţie cu bază mobilă.

Prin logaritmare, relaţia (22.18) devine:

$$lg\,\bar{I}_y = \frac{1}{n}(lg\,y_n - lg\,y_0)$$

şi

$$\bar{I}_y = anti\,lg\left[\frac{1}{n}(lg\,y_n - lg\,y_0)\right]$$

Pe baza datelor din tabelul 22.2. indicele mediu anual de creştere a producţiei de energie electrică în perioada 1950-1989 se calculează astfel:

$$lg\,\bar{I}_{y_{1950-1989}} = \frac{1}{39}(lg\,75851 - lg\,21130 = \frac{1}{39}(4{,}87996 - 3{,}32490) = 0{,}03987$$

Scanned with OKEN Scanner

iar

$$\bar{I} = anti\,lg\,0{,}03987 = 1{,}096$$

Deci producţia de energie electrică a *României* a crescut în perioada 1950-1989 în medie în fiecare an faţă de anul anterior de 1,096 ori.

Metoda mediei geometrice de determinare a indicelui mediu de dezvoltare a fenomenelor, prezintă dezavantajul că ia în calcul numai valorile extreme ale unei serii y_0 şi y_n şi neglijează toţi ceilalţi termeni intermediari.

Din această cauză metoda mediei geometrice se utilizează în practică pentru determinarea indicelui mediu de dezvoltare a fenomenelor numai pentru acele serii dinamice în care indicii variaţiei cu bază mobilă nu diferă în mod substanţial de la o perioadă la alta.

Pentru înlăturarea acestor deficienţe, în literatura de specialitate s-au propus si alte metode, printre care, un loc mai important îl ocupă metoda mediei parabolice şi metoda trendului exponenţial. Aplicarea uneia sau alteia dintre aceste metode se face ţinând seama de proprietăţile lor, de posibilităţile de obţinere a datelor ca şi de esenţa fenomenului studiat.

22.3.1.2. Metoda mediei parabolice

Metoda mediei parabolice de calculare a indicelui mediu înlătură parţial dezavantajele mediei geometrice, luând în calcul suma valorilor termenilor seriei şi valoarea termenului iniţial. În cazul în care termenii unei serii dinamice variază (aproximativ) în progresie geometrică, are loc relaţia:

$$y_0 + y_1 + y_2 + \cdots + y_i + \cdots + y_n = \sum_{i=0}^{n} y_i$$

care prin transformări elementare devine:

$$y_0\bar{I}^0 + y_0\bar{I} + y_0\bar{I}^2 + \cdots + y_0\bar{I}^i + \cdots +_0 \bar{I}^n = \sum_{i=0}^{n} y_i \qquad (22.20)$$

Scanned with OKEN Scanner

$$y_0(1+\bar{I}+\bar{I}^2+\cdots+\bar{I}^i+\cdots+\bar{I}^n)=\sum_{i=0}^{n} y_i$$

$$1+\bar{I}+\bar{I}^2+\cdots+\bar{I}^i+\cdots+\bar{I}^n=\frac{\sum_{i=0}^{n} y_i}{y_0} \qquad (22.21)$$

Primul membru al relaţiei (22.21) reprezintă suma termenilor unei progresii geometrice:

$$1+\bar{I}+\bar{I}^2+\cdots+\bar{I}^i+\cdots+\bar{I}^n=\frac{\bar{I}^{n+1}-1}{\bar{I}-1}$$

deci

$$\frac{\bar{I}^{n+1}-1}{\bar{I}-1}=\frac{\sum_{i=0}^{n} y_i}{y_0} \qquad (22.22)$$

Determinarea indicelui mediu pe baza relaţiei (22.22) se face prin încercări succesive şi de aceea necesită un volum mare de calcule.

În practică, pentru înlăturarea acestei deficienţe, statisticianul chinez *Fin Di Şi* a întocmit un tabel special, pe baza căruia se poate determina indicele mediu, cu ajutorul metodei mediei parabolice, pentru diferite perioade de timp.

Formula (22.22) scoate în evidenţă faptul că metoda mediei parabolice de determinare a indicelui mediu ţine seama de primul termen al seriei dinamice, pentru care se calculează, şi de suma tuturor termenilor ei. Exagerarea importanţei primului termen al seriei reprezintă de fapt şi principala deficienţă a acestei metode.

Scanned with OKEN Scanner

22.3.1.3. Metoda trendului exponenţial

Deficienţele metodei mediei geometrice şi ale metodei mediei parabolice de calcul al indicelui mediu se înlătură, în mare parte, prin utilizarea metodei trendului exponenţial recomandată de literatura de specialitate.

Dacă indicii variaţiei cu bază mobilă sunt apropiaţi ca mărime, atunci variaţia unui fenomen se poate descrie cu ajutorul unei funcţii exponenţiale:

$$y_n = ab^n \tag{22.23}$$

Viteza relativă de creştere a fenomenului y este tocmai parametrul b care joacă rolul de indice mediu de creştere.

Substituind în relaţia (22.23) pe b cu $\bar{I}$ rezultă:

$$y_n = a\bar{I}^n \tag{22.24}$$

Parametrul $\bar{I}$ al funcţiei (22.24) se poate determina uşor cu ajutorul metodei celor mai mici pătrate pe baza formulei:

$$\bar{I} = anti\,lg\frac{\sum_{i=1}^{n} n_i\,lg\,y_i}{\sum_{i=1}^{n} n_i^2} \tag{22.25}$$

Rezultă că, atunci când se cunosc doar termenii extremi ai unei perioade, pentru calculul indicelui mediu se utilizează metoda mediei geometrice, iar dacă se cunosc şi termenii intermediari este recomandabil să se utilizeze metoda mediei parabolice sau metoda trendului exponenţial.

Indicele mediu se poate calcula pentru perioade mai mari sau mai mici de timp. Dacă se cunosc indicii medii de creştere $\bar{I}_j$ ai unui fenomen în diferite subperioade succesive de timp j, atunci pe baza acestora se poate determina indicele

Scanned with OKEN Scanner

mediu de creştere pentru întreaga perioadă $\bar{I}$, ca o medie geometrică ponderată a indicilor de creştere pe subperioade, astfel:

$$\bar{I} = \sqrt[\sum_{j=1}^{k} n_j]{\bar{I}_1^{n_1} \bar{I}_2^{n_2} \dots \bar{I}_j^{n_j} \dots \bar{I}_k^{n_k}} \qquad (22.26)$$

În cazul calculării indicelui mediu al variaţiei pentru întreaga perioadă de timp pe baza unor subperioade neegale între ele se va recurge la formula mediei geometrice ponderate (22.26).

Formula (22.26) permite atât determinarea indicelui mediu de creştere pentru întreaga perioadă, pe baza indicilor medii de creştere ai diferitelor subperioade, cât şi determinarea indicelui mediu al unei subperioade atunci când se cunosc indicii medii de creştere ai celorlalte subperioade şi indicele mediu de creştere pentru întreaga perioadă.

Pe baza indicilor medii ai variaţiei se pot determina indicii medii ai creşterii sau ai sporului cu ajutorul formulei:

$$\bar{r}_y = \bar{I}_y - 1 \qquad sau \qquad \bar{r}_y\% = \bar{I}_y\% - 100$$

Exprimat în procente, indicele mediu al creşterii arată cu cât a crescut în medie în fiecare an al perioadei luate în calcul, nivelul unui anumit fenomen.

22.3.2. Determinarea indicelui mediu pentru fenomene complexe

Metodele de calculare a indicelui mediu de creştere prezentate până acum sunt aplicabile în cazul unor fenomene generale sau singulare, deoarece fac abstracţie de structura lor.

Luând în calcul structura fenomenelor, se pot distinge fenomene complexe care iau naştere prin însumarea unor fenomene simple, iar altele, ca rezultat al produsului mai multor fenomene simple. Astfel, fenomenele complexe pot fi descrise cu ajutorul unor funcţii compuse de forma:

Scanned with OKEN Scanner

$$y = f(X_0, X_1, X_2, \ldots, X_k) = x_0(t) + x_1(t) + \cdots + x_k(t) = \sum_{i=0}^{k} x_i(t) \qquad (22.27)$$

$$y = f(X_0, X_1, X_2, \ldots, X_k) = x_0(t) \cdot x_1(t) \cdots x_k(t) = \prod_{i=0}^{k} x_i(t) \qquad (22.28)$$

De exemplu, producţia lunară Q este egală cu produsul dintre numărul mediu scriptic lunar al personalului T şi productivitatea lunară a muncii Wl, adică:

$$Q = T \cdot Wl .$$

La rândul ei, productivitatea lunară a muncii se poate descompune în alţi doi factori şi anume ca produs între durata medie a lunii de lucru în zile Dz şi productivitatea zilnică Wz astfel:

$$Wl = Dz \cdot Wz$$

iar

$$Q = T \cdot Dz \cdot Wz$$

Productivitea zilnică a muncii se poate descompune în alţi doi factori şi anume durata medie a zilei de lucru, exprimată în ore Dh şi productivitatea orară a muncii Wh:

$$Wz = Dh \cdot Wh$$

şi

$$Wl = Dz \cdot Dh \cdot Wh$$

iar

$$Q = T \cdot Dz \cdot Dh \cdot Wh$$

S-a descompus astfel fenomenul complex - producţie lunară - sub formă de produs a patru factori.

Un alt exemplu, volumul total al desfacerilor de mărfuri este format din suma volumului mărfurilor alimentare, a volumului mărfurilor nealimentare şi a

Scanned with OKEN Scanner

volumului alimentaţiei publice. Volumul consumului de materii prime şi materiale, necesare pentru producerea unui produs, este format din suma tuturor felurilor de materii şi materiale ce se consumă.

În cazul fenomenelor complexe, care reprezintă rezultatul însumării mai multor fenomene simple, indicele mediu al fenomenului complex se va calcula după formula generală (22.18), în care:

$$y_n = x_0(n) + x_1(n) + x_2(n) + \cdots + x_i(n) + \cdots + x_k(n)$$

iar

$$y_0 = x_0(0) + x_1(0) + x_2(0) + \cdots + x_i(0) + \cdots + x_k(0) \qquad i = \overline{0,k}$$

Deci

$$\bar{I}_y = \sqrt[n]{\frac{y_n}{y_0}} = \sqrt[n]{\frac{\sum_{i=0}^{k} x_i(n)}{\sum_{i=0}^{k} x_i(0)}} \qquad (22.29)$$

Formula indicelui mediu (22.29) devine:

$$\sum_{i=0}^{k} x_i(n) = \sum_{i=0}^{k} x_i(0)\bar{I}_{x_i}^{n}$$

$$\bar{I}_y = \sqrt[n]{\frac{\sum_{i=0}^{k} x_i(0)\bar{I}_{x_i}^{n}}{\sum_{i=0}^{k} x_i(0)}} \qquad (22.30)$$

Din relaţia (22.30) rezultă posibilitatea calculării indicelui mediu al unui fenomen complex, pentru o anumită perioadă de timp n, atunci când se cunosc indicii medii ai fenomenelor simple $\bar{I}_{x_i}$ şi nivelurile acestora din perioada de bază $x_i(0)$.

Scanned with OKEN Scanner

Dacă se substituie y_0 şi y_n cu expresiile lor, în formula (22.29) rezultă:

$$\bar{I}_y = \sqrt[n]{\frac{\sum_{i=0}^{k} x_i(n)}{\sum_{i=0}^{k} x_i(0)}} \qquad (22.31)$$

Relaţia (22.31) dă posibilitatea determinării oricărui indice al unui fenomen simplu atunci când se cunoaşte indicele mediu al fenomenului complex pentru aceeaşi perioadă de timp şi indicii medii ai celorlalte fenomene simple. Astfel, să presupunem că se cunosc: indicele mediu al fenomenului complex $\bar{I}_y$, indicii medii ai tuturor fenomenelor simple ce intră în componenţa lui y, cu excepţia lui x_0, nivelurile acestor fenomene din perioada de bază $x_i(0)$. Se cere să se determine indicele mediu al fenomenului x_0, pentru aceeaşi perioadă. Formula generală a indicelui mediu al fenomenului x_0 va fi:

$$\bar{I}_{x_0} = \sqrt[n]{\frac{x_0(n)}{x_0(0)}} \qquad (22.32)$$

Din această formulă nu se cunoaşte elementul $x_0(n)$, a cărui expresie se poate determina după formula:

$$x_0(n) = x_0(o) \cdot \bar{I}_{x_0}^n = y_n - \sum_{i=1}^{k} x_i(n) = y_0 \cdot \bar{I}_y^n - \sum_{i=1}^{k} x_i(o) \cdot \bar{I}_{x_i}^n \qquad (22.33)$$

Substituind pe (22.33) în relaţia (22.32), rezultă:

$$\bar{I}_{x_0} = \sqrt[n]{\frac{y_0 \cdot \bar{I}_y^n - \sum_{i=1}^{k} x_i(o) \cdot \bar{I}_{x_i}^n}{x_0(o)}} \qquad (22.34)$$

sau

Scanned with OKEN Scanner

$$\bar{I}_{x_0} = \sqrt[n]{\frac{y_n - \sum_{i=0}^{k} x_i(n)}{x_0(o)}}$$

În cazul fenomenelor complexe de tipul (22.28) indicele mediu de creştere se determină cu ajutorul formulei:

$$\bar{I}_y = \sqrt{\frac{y_n}{y_0}} = \sqrt[n]{\frac{\prod_{i=0}^{k} x_i(n)}{\prod_{i=0}^{k} x_0(o)}} \qquad i = \overline{0,k} \qquad (22.35)$$

Dezvoltând produsele şi grupând factorii, în ultima expresie a relaţiei (22.35), rezultă:

$$\bar{I}_y = \sqrt[n]{\frac{x_0(n)}{x_0(o)} \cdot \frac{x_1(n)}{x_1(o)} \cdots \frac{x_i(n)}{x_i(o)} \cdots \frac{x_k(n)}{x_k(o)}} \qquad (22.36)$$

dar

$$\frac{x_i(n)}{x_i(0)} = \bar{I}_{x_i} \qquad i = \overline{0,k} \qquad (22.37)$$

$\bar{I}_{x_i}$ = indicele variaţiei cu bază fixă pe întreaga perioadă a fenomenului x_i

Substituind în relaţia (22.36) pe (22.37), se obţine:

$$\bar{I}_y = \sqrt[n]{\prod_{i=0}^{k} \bar{I}_{x_i}} \qquad i = \overline{o,k} \qquad (22.38)$$

Între nivelurile iniţiale şi nivelurile finale ale diferitelor fenomene singulare există relaţia:

Scanned with OKEN Scanner

$$x_i(n) = x_i(o)\bar{I}_{x_i}^n \qquad (22.39)$$

Ţinând seama de relaţia (22.39), indicele mediu al fenomenului complex $\bar{I}_y$, devine:

$$\bar{I}_y = \sqrt[n]{\frac{\prod_{i=0}^{k} x_i(o)\bar{I}_{x_i}}{\prod_{i=0}^{k} x_i(o)}} \quad sau \quad \bar{I}_y = \sqrt[n]{\prod_{i=0}^{k} \bar{I}_{x_i}^n} \qquad (22.40)$$

şi

$$\bar{I}_y = \sqrt[n]{\prod_{i=0}^{k} \bar{I}_{x_i}}$$

Relaţia (22.40) scoate în evidenţă legătura dintre indicele mediu al fenomenului complex şi indicii medii ai fenomenelor simple: indicele mediu al fenomenului complex este egal cu produsul indicilor medii ai fenomenelor simple de tipul (22.28).

22.4. Prognoza fenomenelor pe baza indicatorilor medii de creştere

Prima problemă pe care trebuie să şi-o pună orice cercetător asupra evoluţiei viitoare a unui fenomen este stabilirea pe cale logică a existenţei viitoare a acestuia. Altfel spus, cercetătorul trebuie să stabilească dacă prognoza unui fenomen, într-un orizont dat, este sau nu posibilă. Prognoza unui fenomen poate fi întotdeauna efectuată când există o continuitate în evoluţia lui, când există deci o legătură între nivelul prezent şi cel viitor, când evoluţia din trecut determină într-o anumită măsură, evoluţia lui prezentă şi viitoare, când în evoluţia unui fenomen putem să descoperim o anumită tendinţă de dezvoltare. Dimpotrivă, dacă situaţia unui fenomen s-ar prezenta astfel încât evoluţia lui viitoare ar fi în totalitate ruptă, independentă de prezent şi trecut, să nu decurgă din ceea ce este şi ceea ce a fost, atunci prognoza lui nu ar fi posibilă. Viitorul fenomenelor sociale şi economice se bazează pe modelul lor de evoluţie şi, uneori, pe

Scanned with OKEN Scanner

ipoteza că în decursul orizontului de prognoză considerat, aceste fenomene nu vor fi supuse unor modificări esenţiale.

Două dintre modelele statistice clasice de prognoză mai simple şi mai des utilizate în practică, se fundamentează pe ipoteza variaţiei fenomenelor în progresie aritmetică şi în progresie geometrică.

Ţinând seama de aceste două ipoteze se disting două categorii de metode de prognoză a evoluţiei fenomenelor sociale şi economice:

- metode de prognoză bazate pe sporul mediu;
- metode de prognoză bazate pe indicele mediu de creştere.

22.4.1. Metode de prognoză bazate pe sporul mediu

Evoluţia unui fenomen în progresie aritmetică se realizează atunci când sporurile anuale cu bază mobilă sunt apropiate ca mărime între ele, iar forma evoluţiei lor este liniară.

În condiţiile evoluţiei uniforme, liniare a unui fenomen, după modelul unei progresii aritmetice, se pot rezolva mai multe probleme de prognoză:

22.4.1.1. Determinarea nivelului unui fenomen peste t ani

Să presupunem că un anumit fenomen y a înregistrat, într-o perioadă anterioară, o creştere absolută relativ constantă, dată de un spor mediu egal cu $\overline{\Delta}_y$. Se cere să se determine care va fi nivelul fenomenului y peste n ani.

În ipoteza că şi în perspectivă fenomenul y se va dezvolta tot în progresie aritmetică, cu aceeaşi raţie $\overline{\Delta}_y$, valorile pe care le va atinge în următorii ani vor fi date de relaţia:

$$y_n = y_0 + n\overline{\Delta}_y \qquad (22.41)$$

Scanned with OKEN Scanner

22.4.1.2. Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit

Să considerăm că, pentru a se ajunge la îndeplinirea unor programe de dezvoltare economică şi socială a unei zone dintr-o ţară, este necesar să se dubleze volumul producţiei unui anumit produs de bază y. Cunoscând sporul mediu anual de creştere $\overline{\Delta}_y$ al fenomenului y, se cere să se determine perioada de timp t necesară pentru dublarea fenomenului considerat $y_n = 2y_k$ iar $n-k=t$. Calculele de prognoză se vor efectua în ipoteza că şi în viitor fenomenul y va evolua după aceeaşi tendinţă centrală, dată de modelul progresiei aritmetice, observată în perioada anterioară.

$$y_n = y_k + (n-k)\overline{\Delta}_y \tag{22.42}$$

Substituind pe $y_n = 2y_k$ şi $n-k=t$, în relaţia (22.42), rezultă:

$$2y_k = y_k + t\overline{\Delta}_y$$

$$y_k = t\overline{\Delta}_k$$

de unde

$$t = \frac{y_k}{\overline{\Delta}_k} \tag{22.43}$$

Din relaţia (22.43) rezultă că timpul necesar dublării nivelului unui anumit fenomen este egal cu raportul dintre nivelul prezent al fenomenului respectiv şi sporul lui mediu anual.

22.4.1.3. Stabilirea perioadei de timp t necesară pentru egalizarea nivelului de dezvoltare a două zone A şi B

Să presupunem că o zonă (localitate, judeţ, regiune etc.) A îşi propune să ajungă din urmă o altă zonă B mai dezvoltată din punct de vedere al unui indicator y. Dintr-o perioadă anterioară se cunoaşte sporul mediu anual de creştere a lui y, în fiecare zonă, respectiv $\overline{\Delta}_y^A$ şi $\overline{\Delta}_y^B$ şi se cere să se determine peste câţi ani t zona A

Scanned with OKEN Scanner

ar putea ajunge din urmă zona B. La data efectuării prognozei nivelul indicatorului y în cele două zone era y_0^A şi y_0^B.

Această problemă are sens numai dacă dezvoltarea prezentă şi viitoare a zonei A este şi va fi mai rapidă decât a zonei B.

În momentul egalizării, nivelul lui y, în cele două zone va fi:

$$y_t^A = y_0^A + t\overline{\Delta}_y^A$$

şi

$$y_t^B = y_0^B + t\overline{\Delta}_y^B$$

deci:

$$y_0^A + t\overline{\Delta}_y^A = y_0^B + t\overline{\Delta}_y^B$$

de unde:

$$t(\overline{\Delta}_y^A - \overline{\Delta}_y^B) = y_0^B - y_0^A$$

iar

$$t = \frac{y_0^B - y_0^A}{\overline{\Delta}_y^A - \overline{\Delta}_y^B} \qquad (22.44)$$

Aceasta înseamnă că perioada de timp necesară egalizării nivelului de dezvoltare a celor două zone este egală cu raportul dintre diferenţa între nivelurile de dezvoltare a zonelor în momentul calculului şi diferenţa sporurilor medii ale celor două fenomene obţinute pentru perioada anterioară.

22.4.1.4. Egalizarea nivelului de dezvoltare a două zone peste o perioadă de timp prestabilită

Să presupunem că o zonă A, mai puţin dezvoltată, îşi propune să ajungă din urmă o zonă B, mai dezvoltată în privinţa indicatorului y într-o perioadă de timp prestabilită t. Se cere să se determine sporul mediu anual de creştere $\overline{\Delta}_y^A$ al indicatorului y în zona A necesar pentru asigurarea egalizării cu zona B într-un timp t.

Ca şi în cazul anterior, în momentul egalizării celor două zone se va verifica relaţia:

$$y_0^A + t\overline{\Delta}_y^A = y_0^B + t\overline{\Delta}_y^B$$

Scanned with OKEN Scanner

Presupunând că zona B îşi va păstra în viitor acelaşi spor mediu anual de creştere, atunci din această relaţie va trebui determinat $\overline{\Delta}_y^A$, ceilalţi termeni fiind cunoscuţi:

$$t\,\overline{\Delta}_y^A = y_0^B - y_0^A + t\,\overline{\Delta}_y^B$$

de unde

$$\overline{\Delta}_y^A = \frac{y_0^B - y_0^A}{t} + \overline{\Delta}_y^B, \qquad (22.45)$$

Subliniem faptul că toate prognozele fenomenelor cu ajutorul metodei sporului mediu se bazează pe ipoteza că şi în perspectivă fenomenele luate în calcul se vor dezvolta tot în progresie aritmetică.

22.4.2. Metode de prognoză bazate pe indicele mediu al variaţiei

În ipoteza evoluţiei unui fenomen după modelul unei progresii geometrice prognoza lui se poate face cu ajutorul unor metode bazate pe indicele mediu de creştere, care reprezintă de fapt o estimare a raţiei acestei progresii.

Evoluţia unui fenomen după modelul unei progresii geometrice are loc atunci când indicii de variaţie cu bază mobilă sunt apropiaţi ca mărime între ei.

22.4.2.1. Determinarea nivelului unui fenomen peste t ani

Considerând că fenomenul y s-a dezvoltat în perioada anterioară cu un indice mediu anual de creştere $\overline{I}_y$ şi în momentul efectuării prognozei are nivelul y_0, se cere să se determine care va fi nivelul acestui fenomen peste t ani.

Creşterea lui y în progresie geometrică se exprimă astfel:

$$y_t = y_0\,\overline{I}_y^t \qquad (22.46)$$

Scanned with OKEN Scanner

22.4.2.2. Stabilirea orizontului prognozei în care un anumit fenomen va atinge un nivel prestabilit

Să presupunem că se cere să se determine perioada de timp necesară dublării fenomenului y. În condiţiile notaţiilor anterioare în momentul dublării fenomenului y, se va verifica relaţia:

$$y_t = y_0 \bar{I}_y^t$$

dar

$$y_t = 2y_0 \qquad \text{iar} \qquad 2 = \bar{I}_y^t$$

şi

$$lg\,2 = t\,lg\,\bar{I}_y$$

de unde

$$t = \frac{lg\,2}{lg\,\bar{I}_y} \tag{22.47}$$

22.4.2.3. Stabilirea perioadei de timp necesară pentru egalizarea nivelului de dezvoltare a două zone

Să presupunem că o zonă A, mai puţin dezvoltată în privinţa indicatorului y, îşi propune în viitor să ajungă din urmă o zonă B mai dezvoltată. Pe baza analizei evoluţiei celor două zone, în trecut, se cunosc indicii lor medii de dezvoltare $\bar{I}_y^A$, respectiv $\bar{I}_y^B$ iar nivelurile fenomenului considerat din momentul prognozei erau y_0^A şi respectiv y_0^B.

Se cere să se determine, în condiţiile evoluţiei celor două fenomene şi în viitor în progresie geometrică cu aceleaşi raţii, peste câţi ani t zona A va ajunge din urmă zona B.

Problema are sens numai dacă indicele mediu de creştere în zona A este mai mare decât cel din zona B.

Pe baza relaţiei (22.46) în momentul egalizării va avea loc egalitatea:

Scanned with OKEN Scanner

$$y_0^A \cdot \bar{I}_A^t = y_0^B \cdot \bar{I}_B^t$$

Din această relaţie trebuie determinat t :

$$lg\, y_0^A + t\, lg\, \bar{I}_A = lg\, y_0^B + t\, lg\, \bar{I}_B$$

iar

$$t(lg\, \bar{I}_A - lg\, \bar{I}_B) = lg\, y_0^B - lg\, y_0^A$$

de unde

$$t = \frac{lg\, y_0^B - lg\, y_0^A}{lg\, \bar{I}_A - lg\, \bar{I}_B} \qquad (22.48)$$

22.4.2.4. Egalizarea nivelului de dezvoltare a două zone peste o perioadă de timp prestabilită

Să considerăm că o zonă A îşi propune să ajungă din urmă o zonă B mai dezvoltată în privinţa indicatorului y într-o perioadă de timp dinainte stabilită t .

În ipoteza că zona B îşi va păstra acelaşi indice mediu anual de creştere $\bar{I}_B$ - estimat din evoluţia trecută - se cere să se determine cu ce indice mediu anual trebuie să se dezvolte zona A în aşa fel încât după t ani să se producă egalizarea între ele.

În momentul egalizării se va verifica relaţia:

$$y_0^A \cdot \bar{I}_A^t = y_0^B \cdot \bar{I}_B^t$$

Din această relaţie trebuie determinat $\bar{I}_A$:

$$t\, lg\, \bar{I}_A = lg\, y_0^B - lg\, y_0^A + t\, lg\, \bar{I}_B$$

de unde

$$lg\, \bar{I}_A = \frac{lg\, y_0^B - lg\, y_0^A}{t} + lg\, \bar{I}_B \qquad (22.49)$$

Scanned with OKEN Scanner

Metodele de prognoză bazate pe sporul mediu şi pe indicele mediu de creştere au o largă utilizare practică datorită operativităţii rezultatelor, simplităţii calculelor şi gradului lor de precizie satisfăcător. Aceste metode se bazează pe unele generalizări cu caracter de legitate de dezvoltare, obţinute pe baza unor analize multilaterale ale fenomenelor economice şi sociale cercetate şi se utilizează în practică pentru fenomene omogene a căror structură prezintă o anumită stabilitate în timp şi spaţiu.

Scanned with OKEN Scanner

23. TENDINŢA GENERALĂ A FENOMENELOR ÎN TIMP

23.1. Componentele principale ale evoluţiei fenomenelor în timp

Evoluţia fenomenelor în timp este supusă acţiunii unui complex de factori în rândul cărora se pot evidenţia câteva categorii mai importante şi anume: *factori esenţiali*, care ţin de natura internă a fenomenelor, *factori întâmplători*, accidentali, *factori cu acţiune sezonieră* şi *factori cu acţiune ciclică*.

În dezvoltarea fenomenelor în timp sub influenţa diferitelor categorii de factori care acţionează asupra lor se pot distinge, separa şi determina mai multe componente:

- *o direcţie fundamentală de evoluţie*, dată de influenţa factorilor esenţiali ai dezvoltării, numită şi *tendinţa de dezvoltare* de lungă durată sau, mai simplu, *trend* T;
- *variaţii sau oscilaţii ciclice* C determinate de factori sociali şi economici;
- *variaţii sezoniere* S determinate de factori naturali cu acţiune sezonieră;
- *variaţii întâmplătoare* A determinate de influenţa unor factori aleatori.

Analiza variaţiei fenomenelor în timp înseamnă identificarea, măsurarea separarea şi determinarea celor patru componente ale seriilor dinamice.

În funcţie de natura factorilor şi de relaţiile dintre ei modelul sintetic pe componente al unei serii dinamice se poate prezenta cu ajutorul *schemei aditive*, a *schemei multiplicative* sau a *schemei mixte*.

Schema aditivă se bazează pe relaţia de independenţă dintre componentele seriei dinamice:

$$Y = T + C + S + A \tag{23.1}$$

sau regrupând trendul T şi componenta ciclică C rezultă:

$$Y = (T + C) + S + A$$

Scanned with OKEN Scanner

şi

$$Y = G + S + A$$

unde $G = T + C$ şi reprezintă *seria desezonalizată.*

Schema multiplicativă, bazată pe o relaţie de proporţionalitate a factorilor cu trendul, este descrisă de modelul:

$$Y = T \cdot C \cdot S \cdot (1 + A) \qquad (23.2)$$

In cazul seriilor la care separarea între trend şi ciclu nu este aplicabilă, modelul (23.2) devine:

$$Y = F \cdot S \cdot (1 + A)$$

unde $F = T \cdot C$ şi reprezintă *seria desezonalizată.*

Schema mixtă sau *aditiv-multiplicativă* este descrisă de relaţia:

$$Y = (T + C) \cdot S \cdot (1 + A) \qquad (23.3)$$

O primă problemă care trebuie rezolvată înainte de aplicarea unei metode de separare a componentelor unei serii dinamice este *stabilirea existenţei fiecăreia din cele patru componente.* Desigur întotdeauna există cel puţin o componentă aleatoare A alături de trendul T care se prezintă sub forma unei stabilităţi pe termen lung.

Dacă indicatorii seriei dinamice sunt anuali, indiferent de natura fenomenului cercetat, atunci nu se poate vorbi de existenţa unei componente sezoniere. În mod corespunzător, dacă seria dinamică se referă la o perioadă scurtă de timp, atunci componenta ciclică se omite.

Pentru verificarea existenţei componentei sezoniere se poate recurge la reprezentarea grafică a variaţiilor sezoniere pentru fiecare grupă anuală de date. *Componenta sezonieră* (dacă există) prezintă o relativă stabilitate an de an. Dacă variaţiile sezoniere sunt proporţionale cu valoarea $F = T \cdot C$ atunci se aplică schema multiplicativă iar dacă variaţiile sezoniere rămân stabile dar independente de valoarea $G = T + C$ se aplică schema aditivă.

Scanned with OKEN Scanner

În practică se recomandă aplicarea metodei clasice bazată pe reprezentarea grafică, în care pe axa absciselor se consideră mediile lunare $\bar{y}$ ale indicatorului y, iar pe ordonată, valorile abaterilor standard σ_t ale variabilei y, lunare pentru anul t (fig. 23.1., fig.23.2.) date de relaţiile:

$$\bar{y}_t = \frac{y_{ian.t} + y_{feb.t} + \cdots + y_{dec.t}}{12}$$

şi

$$\sigma_t = \sqrt{\frac{\sum_{1}^{12}(y_{it} - \bar{y}_t)^2}{12}}$$

unde i reprezintă lunile anului.

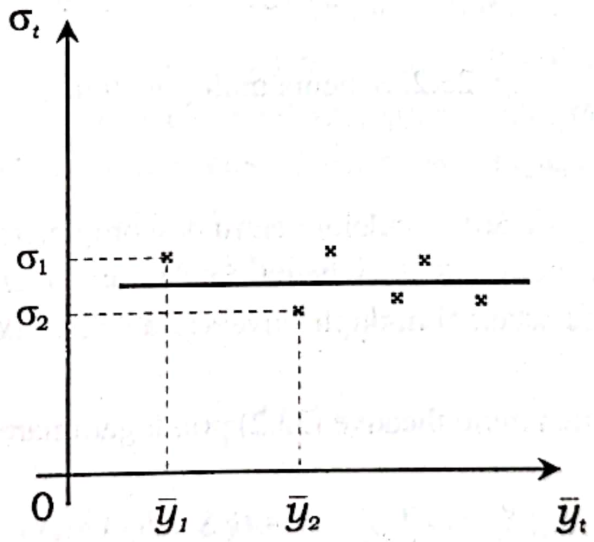


Figura 23.1. Schema aditivă

Dacă punctele (y_t, σ_t) se dispersează aproximativ în jurul unei drepte orizontale, deci $\sigma_t \cong constant$, înseamnă că variaţiile sezoniere sunt stabile şi se poate aplica schema aditivă (fig.23.1).

Scanned with OKEN Scanner

Dacă punctele (y_t, σ_t) se concentrează aproximativ în jurul unei drepte, a cărei înclinaţie faţă de axe este dată de creşterea lui σ_t an de an, atunci variaţiile sezoniere sunt variabile în raport cu tendinţa centrală şi se aplică schema multiplicativă (fig. 23.2).

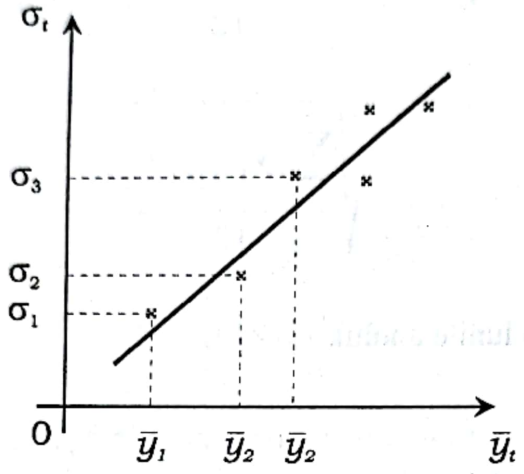


Fig. 23.2. Schema multiplicativă

În vederea simplificării calculelor pentru descompunerea unei serii dinamice în componentele sale, indiferent de schema de descompunere adoptată, se poate recurge la transformarea schemei multiplicative sau a celei mixte în schema aditivă, astfel:

- în cazul schemei multiplicative (23.2) prin logaritmare rezultă:

$$lg\,Y = lg\,T + lg\,C + lg\,S + lg(1 + A) \qquad (23.4)$$

sau

$$Y' = T' + C' + S' + A'$$

unde

$$T' = \lg T;\quad C' = \lg C;\quad S' = \lg S;\quad \text{şi}\quad A' = \lg A$$

- în cazul schemei mixte (23.3) prin logaritmare se obţine:

$$\lg Y = \lg(T + C) + \lg S + \lg(1 + A) \qquad (23.5)$$

sau

$$Y' = G' + S' + A'$$

23.2. Metode de ajustare a seriilor dinamice

Factorii esenţiali determină *tendinţa generală de dezvoltare*, stabilă şi, de regulă, de lungă durată, numită şi *trend*. Separarea componentei de bază a dezvoltării fenomenelor în timp - *trendul* - se obţine prin operaţia de *ajustare*, numită uneori şi *netezire*, a seriilor dinamice.

Determinarea tendinţei centrale de dezvoltare a fenomenelor prezintă o importanţă deosebită pentru cunoaşterea legităţilor lor de dezvoltare, iar aceasta constituie baza de fundamentare ştiinţifică a deciziilor privind evoluţia viitoare a fenomenelor şi proceselor din natură şi societate.

Descoperirea legităţilor de dezvoltare în timp a fenomenelor nu este posibilă decît prin analiza acestora pe *o perioadă suficient de lungă din trecut*, aceasta reprezentând o cerinţă esenţială a acţiunii legii numerelor mari.

A ajusta o serie dinamică (t_i, y_i) înseamnă a înlocui termenii seriei reale cu alţii, obţinuţi cu ajutorul unor metode diferite, şi care reflectă tendinţa generală de lungă durată de dezvoltare a fenomenelor.Termenii seriei ajustate corespund valorilor succesive ale unei funcţii continue $y(t)$.

Timpul joacă rolul de variabilă independentă sau explicativă iar seria dinamică se poate considera ca o funcţie de timp t :

$$y_t = f(t) + \varepsilon_t \qquad (23.6)$$

unde εt reprezintă o parte a indicatorului y_t care nu poate fi explicată pe seama funcţiei $f(t)$ şi se numeşte *reziduu* sau *fluctuaţie aleatoare*.

Mărimea şi sensul termenului ε_t variază în timp odată cu indicatorul y şi depinde numai de întâmplare. Suma tuturor valorilor ε_t, pe întreaga scară de variaţie a timpului, tinde către zero pe măsură ce numărul cazurilor observate (lungimea seriei dinamice) creşte din ce în ce mai mult. Aceasta înseamnă că suma algebrică a abaterilor reziduale $\varepsilon_t = y_t - f(t)$ tinde către zero atunci când numărul termenilor seriei dinamice tinde către infinit. Este clar că funcţia $f(t)$ va fi cu atât mai

Scanned with OKEN Scanner

reprezentativă pentru evoluţia indicatorului y_t cu cât numărul termenilor seriei dinamice, care stă la baza estimării acestei funcţii, este mai mare. Necesitatea luării în calculul tendinţei a unui număr suficient de mare de termeni ai unei serii dinamice este uneori uitată în aplicaţiile practice, fapt care transformă operaţia de ajustare într-un simplu exerciţiu de calcul fără nici o semnificaţie reală.

Pentru estimarea valorilor seriei ajustate se pot aplica metode şi tehnici diferite, în funcţie de natura fenomenului cercetat şi de gradul de precizie urmărit, cum ar fi: metoda grafică, metode mecanice, metode analitice, etc..

Prezentăm în continuare câteva metode dintre cele mai des aplicate în practica statistică.

23.2.1. Metoda grafică

A ajusta o serie dinamică cu ajutorul metodei grafice înseamnă a trasa vizual tendinţa de dezvoltare a fenomenului supus cercetării şi a obţine astfel o primă imagine asupra formei trendului.

Tabelul 23.1. Producţia produsului y în perioada 1975-1990

Anii	Mii tone y_i
1975	130
1976	150
1977	150
1978	190
1979	240
1980	290
1981	420
1982	380
1983	390
1984	350
1985	400
1986	440
1987	445
1988	430
1989	480
1990	500
Total	10210

Date convenţionale

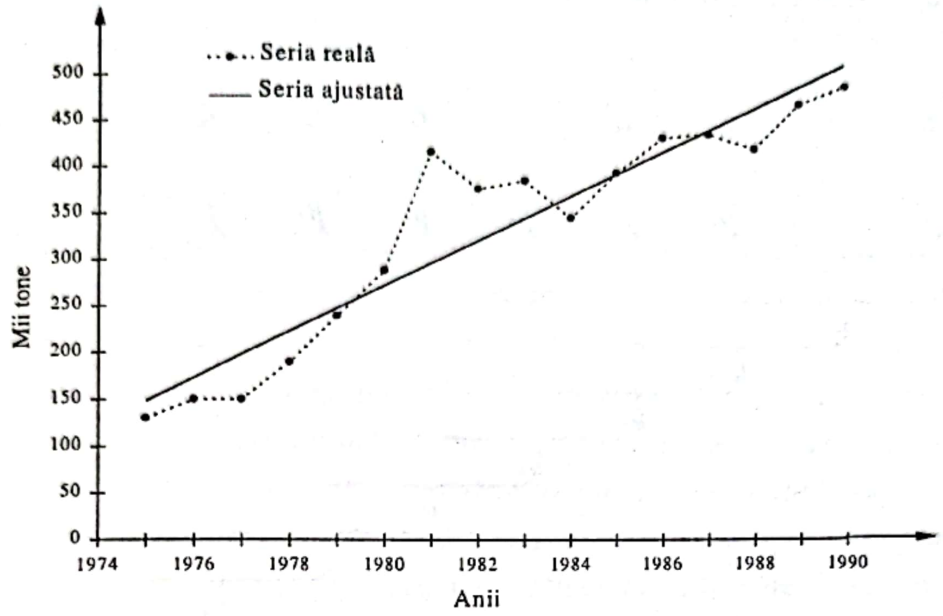


Figura 23.3. Tendinţa liniară determinată prin metoda grafică

În cazul seriilor dinamice care prezintă oscilaţii sezoniere ciclice, pentru a se înlătura variaţiile sezoniere se poate considera, pentru fiecare ciclu, nivelul mediu al indicatorului y utilizându-se în acest scop *metoda mediilor eşalonate*. Ordonatele egale cu mediile eşalonate vor fi ridicate din punctele de abscisă corespunzătoare centrului ciclurilor de variaţie.

23.2.2. Metoda mediilor mobile

Deseori în practica statistică se întâlnesc serii dinamice a căror tendinţă nu poate fi descrisă cu ajutorul unor funcţii simple obişnuite. De exemplu, cazurile în care variaţia ia forma unor curbe trigonometrice (sinusoidă, cosinusoidă, etc.), prezentând anumite oscilaţii ale căror sens şi intensitate se repetă ciclic sau sezonier. În aceste cazuri se poate utiliza metoda mediilor mobile pentru ajustare.

Dacă ciclurile sunt aproximativ egale ca durată, adică cuprind acelaşi număr de termeni, atunci se recomandă ca numărul termenilor din care se calculează mediile mobile să fie egal cu numărul de termeni dintr-un ciclu. Egalitatea duratei ciclurilor

Scanned with OKEN Scanner

este numai uneori realizată în practică şi, de aceea, în cazul general se consideră că numărul termenilor din care se calculează mediile mobile este egal cu numărul termenilor cuprinşi în ciclul de durată medie.

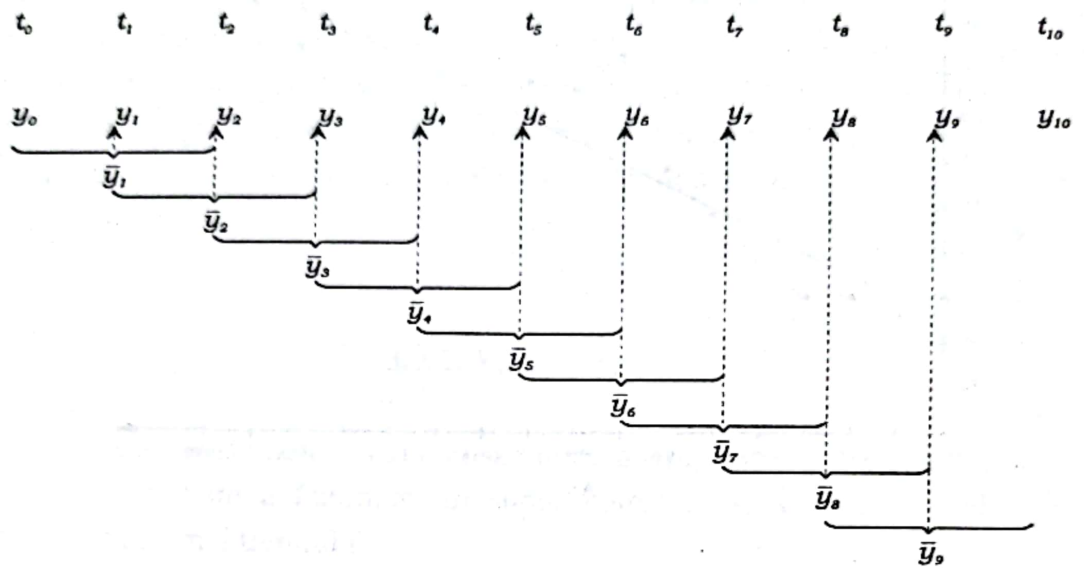


Figura 23.4. Ajustarea unei serii dinamice cu ajutorul metodei mediilor mobile calculate dintr-un număr impar de termeni

Mediile mobile se calculează ca medii aritmetice pe grupe de termeni, obţinute prin deplasarea succesivă în timp cu câte un termen, prin eliminarea primului termen şi includerea termenului următor ultimului luat în calculul mediei anterioare, după formula:

- în momentul t:

$$\bar{y}_t = \frac{y_{t-m} + y_{t-m+1} + \cdots + y_t + \cdots + y_{t+m-1} + y_{t+m}}{2m+1}$$

- în momentul $t+1$:

$$\bar{y}_{t+1} = \frac{y_{t-m+1} + y_{t-m+2} + \cdots + y_{t+1} + \cdots + y_{t+m} + y_{t+m+1}}{2m+1}$$

• în momentul t+2:

$$\bar{y}_{t+2} = \frac{y_{t-m+2} + y_{t-m+3} + \cdots + y_{t+2} + \cdots + y_{t+m+1} + y_{t+m+2}}{2m+1}$$

.....

Aplicarea în practică a acestei metode nu ridică nici un fel de probleme în cazul în care mediile se calculează dintr-un număr impar de termeni : $n = 2m+1$, deoarece fiecare medie mobilă va cădea în centrul intervalului de timp, chiar pe unul din termenii seriei reale.

În cazul calculării mediilor mobile dintr-un număr par de termeni: $n = 2m$, mediile vor cădea între doi termeni centrali ai intervalului respectiv şi va fi nevoie de o nouă operaţie de calcul numită *centrarea mediilor mobile* pentru a determina termenii seriei ajustate, care să cadă chiar în momentele t_i. Centrarea mediilor mobile înseamnă determinarea mediei aritmetice din câte două medii mobile consecutive. Schematic, operaţia de ajustare se prezintă în fig. 23.5.

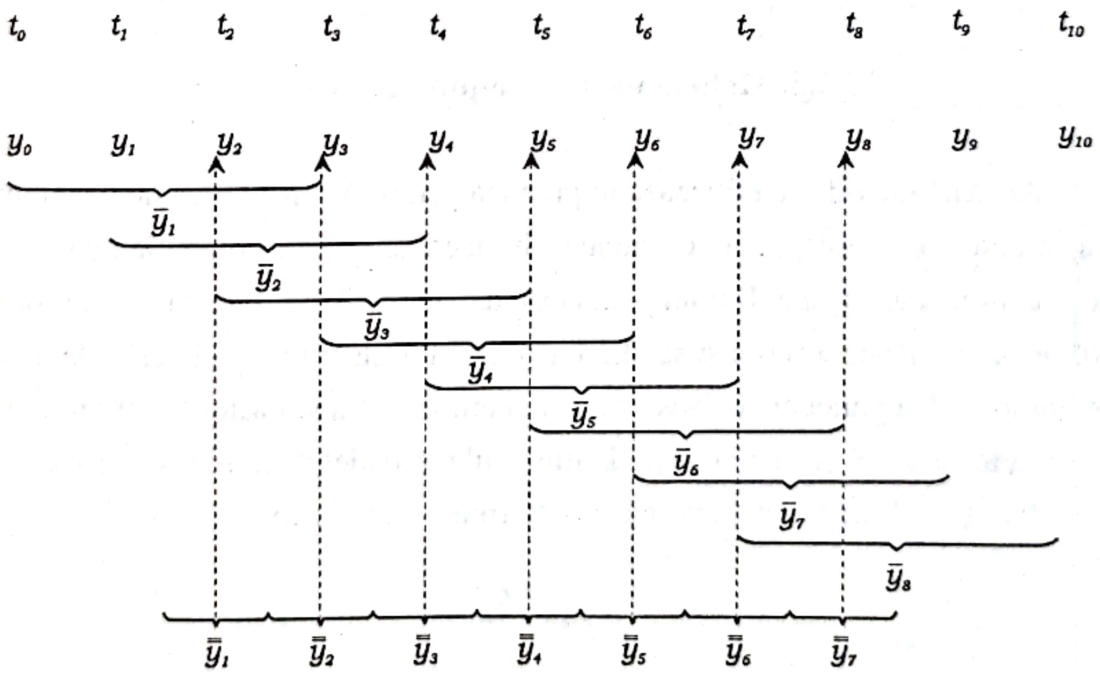


Figura 23.5. Ajustarea unei serii dinamice cu ajutorul metodei mediilor mobile calculate dintr-un număr par de termeni

Scanned with OKEN Scanner

Metoda mediilor mobile, folosită în ajustarea seriilor dinamice, prezintă o serie de *dezavantaje* inerente.

În primul rând, *nu acoperă întreaga perioadă de timp* dintr-o serie dinamică reală. Astfel, dacă numărul termenilor k, din care se calculează mediile mobile, este impar, atunci $(k-1)/2$ valori ale seriei ajustate se pierd la fiecare capăt al seriei reale, iar când k este par, atunci se pierd la fiecare capăt al seriei câte $k/2$ termeni.

În al doilea rând, această metodă *nu duce la obţinerea unui trend propriu-zis*, la o ecuaţie a acestuia, la o lege de dezvoltare, ci doar atenuează, într-o anumită măsură, abaterile maxime şi cele minime ale unei variabile.

În al treilea rând, *mediile mobile simple tind să urmeze datele reale* spre valorile lor maxime şi minime care au cea mai mare amplitudine şi durată.

În sfîrşit, în unele cazuri, *evoluţia obţinută* pe baza aplicării metodei mediilor mobile *se depărtează foarte mult de evoluţia reală*, de exemplu, în cazul tendinţei exponenţiale etc.

Toate aceste neajunsuri ale metodei mediilor mobile de ajustare a unei serii dinamice îi îngustează sfera de aplicabilitate practică.

23.2.3. Metoda mediilor înjumătăţite

Această metodă se utilizează în practică pentru determinarea unei tendinţe de formă liniară care trece prin două puncte. În acest scop seria dinamică empirică se împarte în două părţi egale. Primul punct este dat de media aritmetică a valorilor lui y, pentru prima jumătate a seriei şi se află localizat în mijlocul acestei perioade de timp, iar cel de-al doilea punct are ca abscisă media aritmetică a valorilor lui y din cea de-a doua jumătate a seriei şi se află situat în mijlocul perioadei de timp respective.

Parametrii dreptei de tendinţă se determină pe baza formulelor:

$$a = \frac{2S_1}{n} \tag{23.7}$$

şi

$$b = \frac{4(S_2 - S_1)}{n^2} \tag{23.8}$$

în care

S_1 = suma valorilor lui y din prima jumătate a seriei;
S_2 = suma valorilor lui y din a doua jumătate a seriei;
n = numărul termenilor seriei.

Pentru exemplificare se consideră seria dinamică din tabelul 23.1.

Calculând sumele S_1 şi S_2 şi apoi substituindu-le în formulele (23.7) şi (23.8) rezultă :

$$a = \frac{2S_1}{n} = \frac{2 \cdot 1850}{16} = 231{,}25$$

$$b = \frac{4(S_2 - S_1)}{n^2} = \frac{4(3435 - 1850)}{16^2} = 24{,}77$$

Deci tendinţa liniară de creştere a produsului y va fi descrisă de ecuaţia

$$y_{t_i} = 231{,}25 + 24{,}77 \cdot t_i$$

Tabelul 23.2. Ajustarea unei serii dinamice prin metoda mediilor înjumătăţite

Anii	y_i	S_i	t_i	$y_{t_i} = a + bt_i$
1975	130		-4	144,6
1976	150		-3	169,3
1977	150		-2	194,1
1978	190	S_1=1850	-1	218,9
1979	240		1	243,6
1980	290		2	268,4
1981	320		3	293,2
1982	380		4	317,9
1983	390		5	342,7
1984	350		6	367,5
1985	400		7	392,3
1986	440	S_2=3435	8	417,0
1987	445		9	441,8
1988	430		10	466,6
1989	480		11	491,3
1990	500		12	516,1

Scanned with OKEN Scanner

Pentru a obţine valorile variabilei t_i se consideră $y_0 = a$ în centrul primei jumătăţi a seriei (vezi tabelul 23.2.).

Seria dinamică empirică şi seria ajustată se prezintă în fig. 23.6.

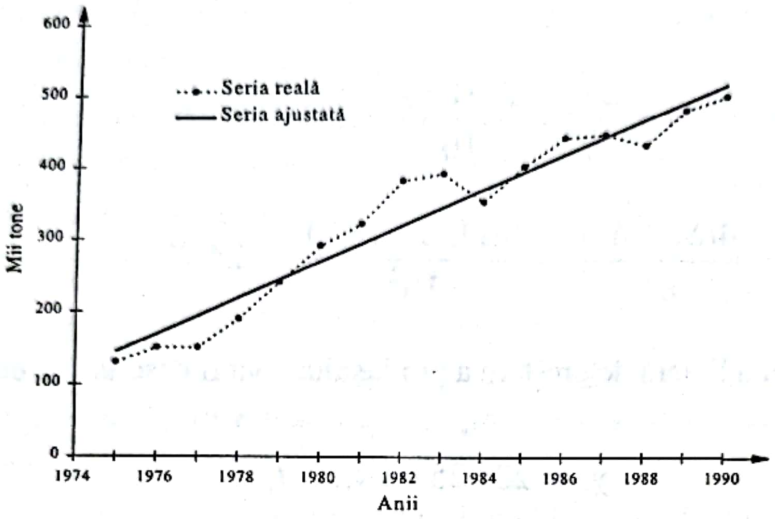


Figura 23.6.Tendinţa centrală obţinută pe baza metodei mediilor înjumătăţite

În cazul unei serii formate dintr-un număr par de termeni $n = 2p+1$, aceasta se poate împărţi în două părţi egale, iar dacă n este impar atunci termenul din centrul seriei se poate cuprinde în ambele părţi sau se poate neglija.

23.2.4. Metoda sporului mediu

Metoda sporului mediu se aplică în practica statistică pentru ajustarea seriilor dinamice ai căror termeni variază aproximativ după modelul unei progresii aritmetice. Raţia progresiei aritmetice se consideră egală cu sporul mediu de creştere a fenomenului calculat pe baza seriei dinamice luate în studiu.

Variaţia unui fenomen în timp se apropie de modelul unei progresii aritmetice dacă sporurile cu bază mobilă $\Delta_{i/i-1}$ sunt apropiate între ele ca mărime. În particular, dacă aceste sporuri sunt chiar egale între ele, atunci dezvoltarea fenomenului cercetat se desfăşoară exact după modelul unei progresii aritmetice, situaţie care se poate întâlni în practică numai în mod cu totul excepţional. Dezvoltarea fenomenelor sociale şi economice reale se apropie într-o măsură mai mare sau mai mică de modelul teoretic al unei progresii aritmetice dar, de regulă, nu se va suprapune peste acest model.

Scanned with OKEN Scanner

Termenii seriei ajustate se determină pe baza formulei termenului general al unei progresii aritmetice:

$$y_k = y_0 + k\overline{\Delta}_y \qquad k = 0,1,2,\ldots,n \tag{23.9}$$

în care:

y_k = un termen oarecare al seriei ajustate;

y_0 = nivelul considerat ca bază;

$\overline{\Delta}_y$ = sporul mediu al lui y.

Deoarece ecuaţia (23.9) este similară cu ecuaţia unei drepte, forma tendinţei, obţinută pe baza metodei sporului mediu, va fi întotdeauna liniară.

Se ştie că, în cazul unei progresii aritmetice, y_0 reprezintă nivelul iniţial, cel mai mic ca valoare. Aceasta însemnă că în cazul ajustării seriilor dinamice crescătoare, după metoda sporului mediu, se poate considera ca nivel de bază primul termen al seriei reale. În acest caz, prin adăugarea în mod succesiv a sporului mediu, se va ajunge în final ca valoarea teoretică a lui y să coincidă cu ultimul termen al seriei reale. Deci, tendinţa va trece în mod obligatoriu prin punctele extreme ale seriei şi va face abstracţie de toate celelalte intermediare, uneori situându-se în totalitate deasupra sau dedesubtul punctelor reale intermediare, diminuând caracterul de reprezentativitate al tendinţei. De aceea, acest procedeu se recomandă a se aplica în practică numai în cazurile în care, fie nu se cunosc decât valorile extreme ale unei serii dinamice, fie că se cunosc şi termenii intermediari, dar nivelurile extreme reprezintă niveluri de dezvoltare normală a fenomenului.

În toate celelate cazuri este necesar să se identifice, prin analiza seriei, care este nivelul ce se apropie în cea mai mare măsură de tendinţa normală de dezvoltare şi apoi să se considere acesta ca nivel de bază y_0. Un procedeu simplu de alegere a nivelului de bază constă în reprezentarea grafică a seriei urmată de trasarea vizuală a tendinţei. Ordonata punctului care se va apropia în cea mai mare măsură de tendinţa astfel trasată se va considera nivel de bază y_0.

Gradul de reprezentativitate a tendinţei unei serii dinamice se poate aprecia cu ajutorul indicatorului :

$$\sum_{i=0}^{n} |y_i - y_t| \qquad i = \overline{0,n}$$

Exemplu. Fie seria dinamică din tabelul 23.1 care exprimă o evoluţie liniară a fenomenului y. Întrucât termenii externi ai seriei se apropie în mare măsură de tendinţa de evoluţie, trasată vizual se poate considera drept origine de calcul nivelul y_0 corespunzător anului 1975. Raţia progresiei aritmetice este egală cu sporul mediu $\overline{\Delta}_y$:

Scanned with OKEN Scanner

$$\overline{\Delta}_y = \frac{y_n - y_0}{n-1} = \frac{500-130}{16-1} = 24{,}66$$

În tabelul 23.3 se prezintă ajustarea unei serii dinamice cu ajutorul metodei sporului mediu utilizând două modalităţi de stabilire a originii de calcul y_0.

Astfel, în primele 5 coloane ale tabelului se consideră drept *origine de calcul* y_0 corespunzător anului 1975, adică primului termen al seriei. Din coloana 5 se observă că majoritatea termenilor seriei ajustate se abat în plus faţă de valorile lor empirice şi numai o mică parte se abat în minus faţă de tendinţă:

$$\sum_{i=0}^{n} \left| y_i - y_{t_i} \right| = 286{,}66$$

Se încalcă în felul acesta principiul potrivit căruia tendinţa de evoluţie a fenomenelor trebuie să fie o tendinţă medie faţă de care abaterile într-un sens să fie egale în valoare absolută cu abaterile în sens contrar.

Tabelul 23.3 Ajustarea unei serii dinamice cu ajutorul metodei sporului mediu.

Anii	y_i	t_i	$y_{t_i} = y_0 - t_i\overline{\Delta}_y$	$y_i - y_{t_i}$	t_i'	$y'_{t_i} = y'_0 + t'_i\overline{\Delta}_y$	$y_i - y'_{t_i}$
	$y_{1975} = y_0$					$y_{1979} = y_0$	
1975	130	0	130,00	0,00	-4	141,36	-11,36
1976	150	1	154,66	-4,66	-3	166,02	-16,02
1977	150	2	179,32	-29,32	-2	190,68	-40,68
1978	190	3	203,99	-13,99	-1	215,34	-25,34
1979	**240**	4	228,64	11,36	0	240,00	0
1980	290	5	253,30	36,70	1	264,66	25,34
1981	320	6	277,96	42,04	2	289,32	30,68
1982	380	7	302,62	77,38	3	313,98	66,02
1983	390	8	327,28	62,72	4	338,64	51,36
1984	350	9	351,94	-1,94	5	360,30	-13,30
1985	400	10	376,60	23,40	6	387,96	12,04
1986	440	11	401,26	38,74	7	412,62	27,38
1987	445	12	425,92	19,08	8	437,28	7,72
1988	430	13	450,58	20,58	9	461,94	-31,94
1989	480	14	475,24	4,76	10	486,60	-6,60
1990	500	15	499,90	-0,10	11	511,26	-11,26
Total	5285	-	5039,27	286,66		5220,96	64,04

Scanned with OKEN Scanner

Pentru a înlătura în mare măsură această deficienţă a tendinţei determinată de alegerea mecanică a primului termen al seriei, drept origine de calcul se va utiliza cea de-a doua modalitate de stabilire a originii de calcul considerând $y_{1979} = y_0$. Calculul tendinţei s-a efectuat în partea a doua a tabelului 23.3. Se observă că de data aceasta tendinţa de evoluţie este mai apropiată de tendinţa medie: $\sum_{i=0}^{n}|y_i - y_i'|$ este mai mică decât în cazul precedent, diminuându-se abaterile în plus faţă de tendinţă şi crescând substanţial cele în minus: 377,04 < 386,86.

23.2.5. Metoda indicelui mediu de variaţie

Această metodă se poate utiliza în practică pentru ajustarea seriilor dinamice ai căror termeni variază aproximativ după modelul unei progresii geometrice, iar expresia ei grafică ia forma unei curbe exponenţiale. În general, se apreciază că o serie dinamică prezintă variaţia unui fenomen în progresie geometrică dacă indicii variaţiei cu bază mobilă se apropie între ei ca mărime.

Termenii seriei ajustate se determină pe baza formulei termenului general al unei progresii geometrice:

$$y_k = y_0 \cdot \bar{I}_y^k \qquad k = 0,1,2,\ldots,n \tag{23.10}$$

în care: y_k = un termen oarecare al progresiei geometrice;

y_0 = nivelul lui y considerat ca origine de calcul;

$\bar{I}_y$ = indicele mediu de variaţie al seriei dinamice reale;

k = numărul de termeni care separă pe y_k de y_0.

Ca şi în cazul metodei sporului mediu, în aplicarea acestei metode, originea de calcul y_0 se poate considera ca fiind egală cu nivelul iniţial al lui y din cadrul seriei, sau mai bine, se alege pe baza ajustării vizuale a seriei, ca fiind egal cu acel nivel al seriei care se apropie în cea mai mare măsură de tendinţa trasată vizual.

Pentru exemplificare, considerăm evoluţia producţiei de oţel a *României* în perioada 1950-1976 (vezi tabelul 23.4). Reprezentând grafic această serie (fig. 23.7) se observă că producţia de oţel evoluează aproximativ după o curbă exponenţială şi că unor intervale egale de pe axa timpului le corespund creşteri din ce în ce mai mari ale producţiei de oţel.

Indicele mediu anual al creşterii producţiei de oţel în perioada 1950 -1976 va fi:

Scanned with OKEN Scanner

$$\bar{I} = \sqrt[26]{\frac{10733}{555}} = 1{,}121 \text{ ori} \quad \bar{r} = 12{,}1\%$$

Tabelul 23.4. Ajustarea seriei dinamice cu ajutorul metodei indicelui mediu

Anii	y_i	t_i	$y_{t_i} = y_0 \cdot \bar{I}^{t_i}$	$\lvert y_i - y_k \rvert$
1	2	3	4	5
1950	555	0	555	0
1951	644	1	622	22
1952	698	2	697	1
1953	717	3	782	65
1954	628	4	876	248
1955	766	5	982	216
1956	779	6	1101	322
1957	864	7	1235	371
1958	934	8	1384	450
1959	1420	9	1551	131
1960	1806	10	1739	67
1961	2126	11	1950	176
1962	2451	12	2186	265
1963	2704	13	2450	254
1964	3039	14	2746	293
1965	3426	15	3078	348
1966	3670	16	3451	219
1967	4088	17	3869	219
1968	4751	18	4337	414
1969	5540	19	4862	678
1970	6517	20	5450	1067
1971	6803	21	6110	693
1972	7401	22	6849	552
1973	8161	23	7678	483
1974	8848	24	8607	241
1975	9549	25	9648	99
1976	10733	26	10815	82
Total	99618	-	95610	7976

Sursă: Prelucrat după Anuarele statistice ale României

Alegând drept origine de calcul primul termen al seriei $y_0 = 555$, corespunzător anului 1950, pe col. 3 a tabelului 23.4, s-au determinat valorile succesive ale lui t_i.

Scanned with OKEN Scanner

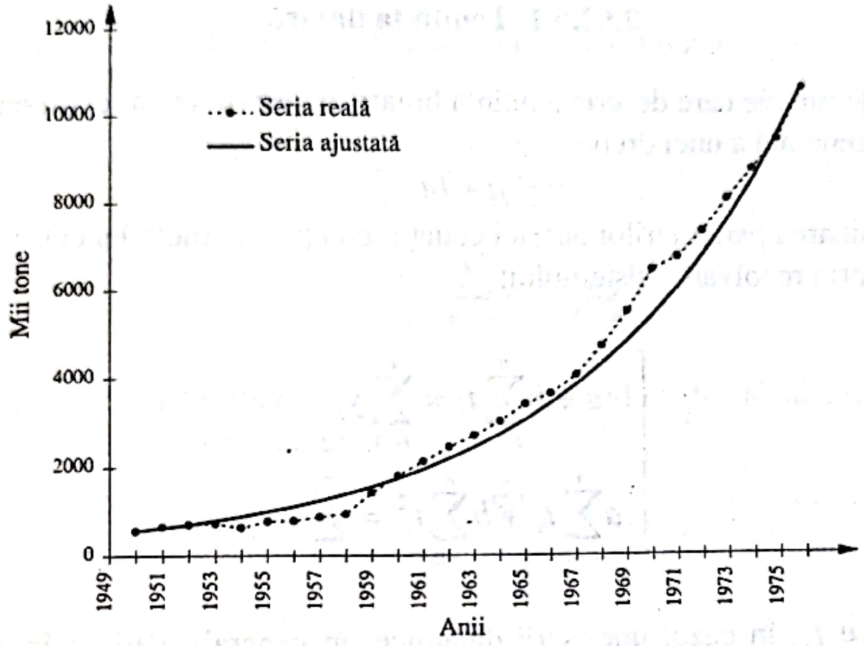


Figura 23.7. Producţia de oţel a *României* în perioada 1950-1976

Substituind în formula (23.10) indicele mediu şi t_i, rezultă termenii seriei ajustate din col. 4 a tabelului 23.4.

23.2.6. Metoda celor mai mici pătrate

Problema fundamentală care trebuie rezolvată în ajustarea unei serii dinamice, folosind această metodă, este alegerea tipului celui mai potrivit de funcţie $f(t)$, pe baza analizei fenomenului luat în calcul şi în concordanţă cu natura lui.

Funcţiile mai des utilizate în practica statistică pentru determinarea tendinţei generale de dezvoltare a fenomenelor sunt: funcţii polinomiale, hiperbole, parabole, exponenţiale etc.

Scanned with OKEN Scanner

23.2.6.1. Tendinţa liniară

Tipul de funcţie care descrie tendinţa liniară de dezvoltare a fenomenelor este dat de ecuaţia generală a unei drepte:

$$y = a + bt$$

Determinarea parametrilor acestei ecuaţii, cu ajutorul metodei celor mai mici pătrate, se face prin rezolvarea sistemului:

$$\begin{cases} na + b\sum_{i=0}^{n} t_i = \sum_{i=0}^{n} y_i \\ a\sum_{i=0}^{n} t_i + b\sum_{i=0}^{n} t_i^2 = \sum_{i=0}^{n} t_i y_i \end{cases} \qquad (23.11)$$

Deoarece t_i, în cazul unei serii dinamice (în general), variază în progresie aritmetică cu raţia egală cu 1, se poate face o mişcare de *translaţie* a dreptei faţă de originea axelor de coordonate în aşa fel încât centrul intervalului de timp luat în calcul să coincidă cu originea axelor de coordonate (fig.23.8).

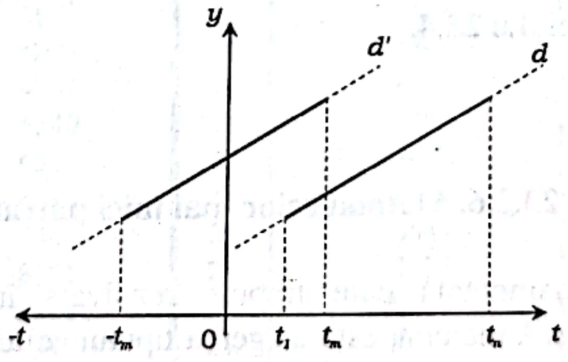


Figura 23.8. Mişcarea de translaţie a dreptei faţă de origine

Abscisele diferitelor puncte ale dreptei d_i se exprimă prin abateri faţă de anul din centrul seriei, considerat drept origine de calcul. Întrucât abscisa punctului din centrul seriei *zeroul fals* este egală cu nivelul mediu al variabilei independente, rezultă, pe baza proprietăţilor mediei aritmetice, că suma abaterilor pozitive este egală în valoare absolută cu suma abaterilor negative şi că ele se compensează reciproc la nivelul seriei:

Scanned with OKEN Scanner

$$\sum_{i=-m}^{m} t_i = 0 \tag{23.12}$$

Substituind relaţia (23.12) în sistemul (23.11) rezultă:

$$\begin{cases} na = \sum_{i=-m}^{m} y_i \\ b\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} t_i y_i \end{cases} \tag{23.13}$$

Rezolvând sistemul (23.13) se obţin formulele de calcul simplificat ale celor doi parametri ai ecuaţiei liniare:

$$a = \frac{\sum_{i=-m}^{m} y_i}{n} \tag{23.14}$$

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2} \tag{23.15}$$

Aplicarea în practică a formulelor (23.14) şi (23.15) se justifică numai în cazul îndeplinirii relaţiei (23.12). Alegerea originii de calcul se face în mod diferit după cum numărul termenilor seriei ce urmează a fi ajustată este par sau impar.

În cazul unei serii formate dintr-un număr impar de termeni, originea de calcul va coincide chiar cu termenul din centrul seriei, iar abaterile faţă de origine se vor exprima în ani.

Exemplu. Se consideră din nou seria dinamică din tabelul 23.1. a cărei tendinţă se apropie de o dreaptă. Calculele intermediare necesare ajustării se prezintă sistematizat în tabelul 23.5.

Aplicând formulele (23.14) şi (23.15) rezultă:

$$a = \frac{\sum_{i=-m}^{m} y_i}{n} = \frac{5285}{16} = 330{,}31 \text{ mii tone}$$

Scanned with OKEN Scanner

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2} = \frac{17265}{1360} = 12{,}69 \text{ mii tone}$$

$a = \bar{y} = 330{,}31$ mii tone arată că producţia medie a produsului y în perioada 1975 - 1990 a fost de 330,31 mii tone care s-a realizat la mijlocul anului 1963 iar $b = 12{,}69$ mii tone înseamnă că în medie în fiecare an al perioadei producţia produsului considerat a crescut cu $2b = 25{,}38$ mii tone.

Tabelul 23.5. Determinarea tendinţei liniare cu ajutorul MCMMP

Anii	y_i	t_i	t_i^2	$t_i y_i$	$y_{t_i} = a + bt_i$	$y_i - y_{t_i}$
1975	130	-15	225	-1950	139,96	-9,96
1976	150	-13	169	-1950	165,34	-15,34
1977	150	-11	121	-1650	190,72	-40,72
1978	190	-9	81	-1710	216,10	-26,10
1979	240	-7	49	-1680	241,48	-1,48
1980	290	-5	25	-1450	266,86	23,14
1981	320	-3	9	-960	292,24	27,76
1982	380	-1	1	-380	317,62	62,38
1983	390	1	1	390	343,00	47,00
1984	350	3	9	1050	368,38	-18,38
1985	400	5	25	2000	393,76	6,24
1986	440	7	49	3080	419,14	20,86
1987	445	9	81	4005	444,52	0,48
1988	430	11	121	4730	469,90	-39,90
1989	480	13	169	6240	495,28	-15,28
1990	500	15	225	7500	520,66	-20,66
Total	5285	0	1360	17265	5284,96	0

Menţionăm faptul că, în cazul unei serii dinamice anuale formată dintr-un număr par de termeni, centrul seriei este situat la mijlocul intervalului de timp dintre cei doi termeni centrali şi se exprimă în jumătăţi de an, iar b va reprezenta sporul mediu pe o jumătate de an.

Scanned with OKEN Scanner

Valorile reale şi cele teoretice ale producţiei produsului y în perioada 1975-1990 se prezintă în fig. 23.9.

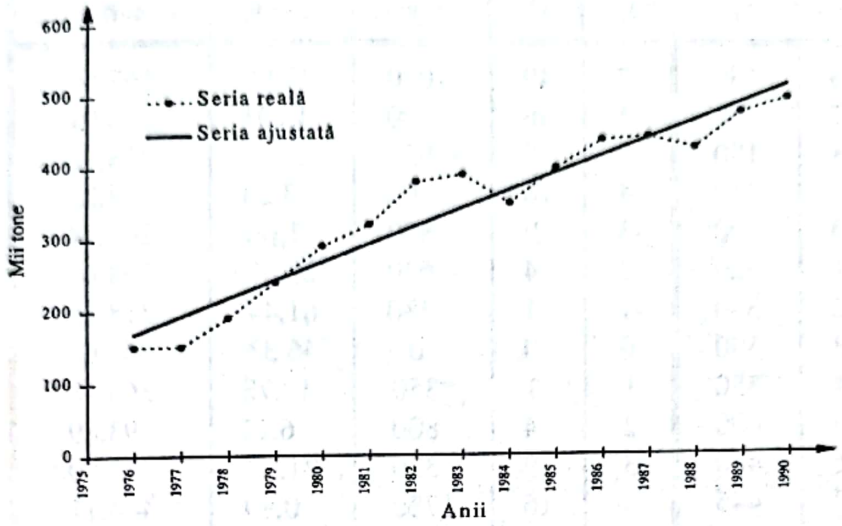


Figura 23.9. Evoluţia producţiei produsului y în perioada 1975-1990

Înlocuind datele în formulele (23.14) şi (23.15) rezultă:

$$a = \frac{\sum_{i=-m}^{m} y_i}{n} = \frac{5155}{15} = 343{,}67 \quad \text{mii tone}$$

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2} = \frac{7030}{280} = 25{,}11 \quad \text{mii tone}$$

Pentru seriile formate dintr-un număr par de termeni, centrul intervalului de timp luat în calcul va coincide cu mijlocul intervalului de timp dintre cei doi termeni centrali ai seriilor. Dacă se va scade din fiecare an originea de calcul, atunci abaterile faţă de origine, din diferiţi ani se vor exprima în jumătăţi de an (în cazul general, în care indicatorii y_i sunt anuali).

Scanned with OKEN Scanner

Tabelul 23.6. Determinarea tendinţei liniare cu ajutorul *MCMMP* în cazul $n = 2m + 1$

Anii	y_i	t_i	t_i^2	$t_i y_i$	$y_i - y_{t_i}$	$y_{t_i} = a + bt_i$
1976	150	-7	49	-1050	-17,90	167,90
1977	150	-6	36	- 900	-43,01	193,01
1978	190	-5	25	- 950	-28,12	218,12
1979	240	-4	16	- 960	- 3,23	243,23
1980	290	-3	9	- 870	21,66	268,34
1981	320	-2	4	- 640	26,55	293,45
1982	380	-1	1	- 380	61,44	318,56
1983	390	0	0	0	46,33	343,67
1984	350	1	1	350	-18,78	368,78
1985	400	2	4	800	6,11	393,89
1986	440	3	9	1320	21,00	419,00
1987	445	4	16	1780	0,89	444,11
1988	430	5	25	2150	-39,22	469,22
1989	480	6	36	2880	-14,33	494,33
1990	500	7	49	3500	-19,44	519,44
Total	5155	0	280	7030	0	5155,05

În cazul tendinţei liniare, determinate cu ajutorul metodei celor mai mici pătrate, *media valorilor teoretice* $\bar{y}_t$ *este egală cu media valorilor empirice* $\bar{y}$.

Valorile teoretice se determină pe baza relaţiei:

$$y_t = a + bt_i$$

iar

$$\sum_{i=0}^{n} y_{t_i} = \sum_{i=0}^{n} (a + bt_i) = na + b\sum_{i=0}^{n} t_i = \sum_{i=0}^{n} y_i$$

deci

$$\sum_{i=0}^{n} y_{t_i} = \sum_{i=0}^{n} y_i \quad \text{şi} \quad \bar{y}_{t_i} = \bar{y}$$

Pe baza acestei relaţii se poate verifica exactitatea calculelor de ajustare în practică.

Scanned with OKEN Scanner

23.2.6.2. Tendinţa parabolică de gradul doi

Aplicarea metodei celor mai mici pătrate pentru estimarea parametrilor parabolei de gradul doi duce la obţinerea următorului sistem de ecuaţii normale:

$$\begin{cases} na + b\sum_{i=0}^{n} t_i + c\sum_{i=0}^{n} t_i^2 = \sum_{i=0}^{n} y_i \\ a\sum_{i=0}^{n} t_i + b\sum_{i=0}^{n} t_i^2 + c\sum_{i=0}^{n} t_i^3 = \sum_{i=0}^{n} t_i y_i \\ a\sum_{i=0}^{n} t_i^2 + b\sum_{i=0}^{n} t_i^3 + c\sum_{i=0}^{n} t_i^4 = \sum_{i=0}^{n} t_i^2 y_i \end{cases} \qquad (23.16)$$

Ca şi în cazul tendinţei liniare, calculele parametrilor tendinţei parabolice se pot simplifica făcând substituţiile $t_i = 0;\ t_i^3 = 0$ în sistemul (23.16) care devine:

$$\begin{cases} na + c\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} y_i \\ b\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} t_i y_i \\ a\sum_{i=-m}^{m} t_i^2 + c\sum_{i=-m}^{m} t_i^4 = \sum_{i=-m}^{m} t_i^2 y_i \end{cases} \qquad (23.17)$$

Din ecuaţia a doua a sistemului (23.17) rezultă:

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2} \qquad (23.18)$$

Rezolvând sistemul (23.17), din prima şi a treia ecuaţie rezultă formulele de calcul ale parametrilor a şi c:

Scanned with OKEN Scanner

$$a = \frac{\sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^2 y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} \qquad (23.19)$$

şi

$$c = \frac{n\sum_{i=-m}^{m} t_i^2 y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} \qquad (23.20)$$

Alegerea originii de calcul în aşa fel încât $\sum_{i=-m}^{m} t_i = 0$ şi $\sum_{i=-mi}^{m} t_i^3 = 0$ se face în acelaşi fel ca şi în cazul tendinţei liniare.

Modelul de trend descris de parabola de gradul doi este aplicabil în practică atunci când diferenţele de ordinul doi ale valorilor empirice y_i sunt apropiate între ele ca mărime. În cazul ideal în care valorile empirice y_i se suprapun peste valorile lor teoretice y_{t_i}, atunci diferenţele de ordinul doi sunt egale cu o constantă.

Pentru exemplificare, să considerăm evoluţia populaţiei municipiului *Iaşi* în perioada 1950-1976 (tabelul 23.7).

Înlocuind valorile din tabel în formulele (23.18), (23.19) şi (23.20), rezultă valorile parametrilor a, b şi c, astfel:

$$a = \frac{\sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^2 y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} = \frac{178542 \cdot 3925 - 1638 \cdot 252207}{27 \cdot 178542 - 1638^2} = 134{,}5732 \text{ mii loc.}$$

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2} = \frac{7587}{1638} = 4{,}6319 \text{ mii loc.}$$

Scanned with OKEN Scanner

$$c = \frac{n\sum_{i=-m}^{m} t_i^2 y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} = \frac{27 \cdot 252207 - 1638 \cdot 3925}{27 \cdot 178542 - 1638^2} = 0{,}1780 \text{ mii loc.}$$

Aceasta înseamnă că evoluţia populaţiei municipiului *Iaşi*, în perioada 1950 - 1976, s-a produs după modelul unei parabole de gradul doi:

$$y_i = 134{,}57 + 4{,}63t_i + 0{,}18t_i^2$$

Dacă substituim succesiv, în această formulă, valorile lui t_i din col. 3 a tabelului 23.7, se obţin valorile modelului parabolic din col. 8 a aceluiaşi tabel.

Pentru a determina numărul populaţiei viitoare a municipiului *Iaşi* se substituie, în ecuaţia de tendinţă, valorile t_i pe care le va lua în anii viitori, astfel:

- în anul 1980: $t_{1980} = 17$; $y_{1980} = 265{,}3$ mii loc.
- în anul 1985: $t_{1985} = 22$; $y_{1985} = 323{,}55$ mii loc.

Pentru a se obţine o simplificare mai mare a calculelor la însumarea puterilor pare ale lui t_i, în sistemul (23.17) se poate face substituţia:

$$\sum_{i=-m}^{m} t_i^{2k} = 2\sum_{i=0}^{m} t_i^{2k} \qquad (23.21)$$

unde $n = 2m + 1$

Sistemul (23.17) devine:

$$\begin{cases} na + 2c\sum_{i=0}^{m} t_i^2 = \sum_{i=-m}^{m} y_i \\ 2b\sum_{i=0}^{m} t_i^2 = \sum_{i=-m}^{m} t_i y_i \\ 2a\sum_{i=0}^{m} t_i^2 + 2c\sum_{i=0}^{m} t_i^4 = \sum_{i=-m}^{m} t_i^2 y_i \end{cases} \qquad (23.22)$$

Scanned with OKEN Scanner

Tabelul 23.7. Evoluţia populaţiei municipiului Iaşi în perioada 1950-1976 şi prognoza până în anul 1985

Anii	(y_i)	t_i	t_i^2	t_i^4	$t_i y_i$	$t_i^2 y_i$	$y_{t_i}=a+bt+ct^2$
1	2	3	4	5	6	7	8
1950	102	-13	169	28561	-1326	17238	104
1951	104	-12	144	20736	-1248	14976	105
1952	105	-11	121	14641	-1155	12705	105
1953	107	-10	100	10000	-1070	10700	106
1954	108	-9	81	6561	-972	8748	107
1955	111	-8	64	4096	-888	7104	109
1956	113	-7	49	2401	-791	5537	111
1957	118	-6	36	1296	-708	4248	113
1958	121	-5	25	625	-605	3025	116
1959	121	-4	16	256	-484	1936	119
1960	122	-3	9	81	-366	1098	122
1961	124	-2	4	16	-240	496	126
1962	125	-1	1	1	-125	125	130
1963	126	0	0	0	0	0	135
1964	126	1	1	1	126	126	139
1965	124	2	4	16	248	496	145
1966	161	3	9	81	483	1449	150
1967	164	4	16	256	656	2624	156
1968	170	5	25	625	850	4250	162
1969	176	6	36	1296	1056	6336	169
1970	182	7	49	2401	1274	8918	176
1971	186	8	64	4096	1488	11904	183
1972	191	9	81	6561	1719	15471	191
1973	198	10	100	10000	1980	19800	199
1974	206	11	121	14641	2266	24926	207
1975	215	12	144	20736	2580	30960	216
1976	219	13	169	28561	2847	37011	225
Total	3925	0	1638	178542	7587	252207	3925
1977	-	14	196	-	-	-	234
1978	-	15	225	-	-	-	244
1979	-	16	256	-	-	-	254
1980	-	17	289	-	-	-	265
1981	-	18	324	-	-	-	276
1982	-	19	361	-	-	-	287
1983	-	20	400	-	-	-	298
1984	-	21	441	-	-	-	310
1985	-	22	484	-	-	-	324

Scanned with OKEN Scanner

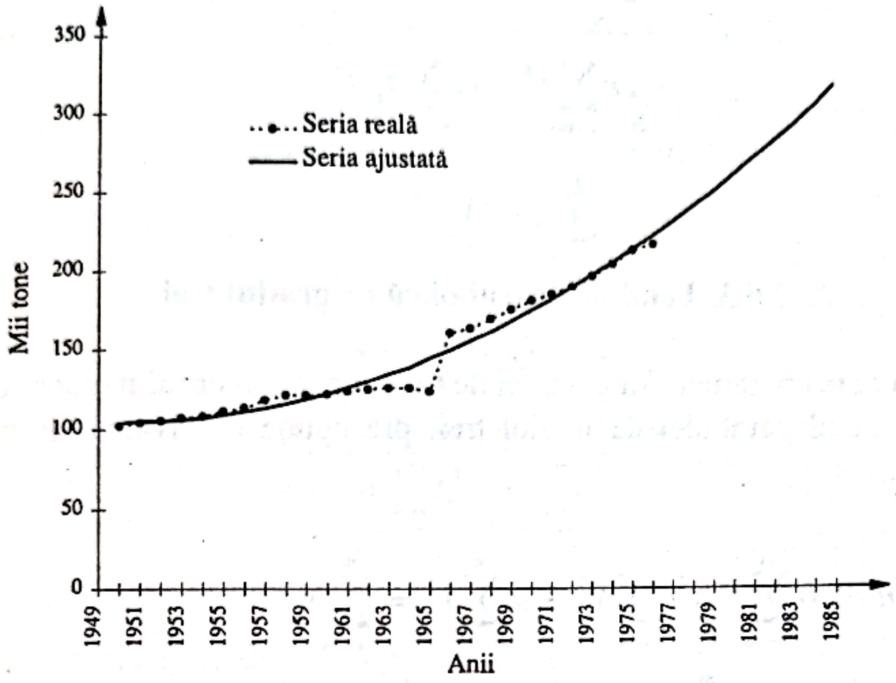


Figura 23.10. Evoluţia populaţiei municipiului *Iaşi* în perioada 1950-1976

Rezolvând sistemul (23.22) rezultă formulele parametrilor parabolei de gradul doi cu ajutorul cărora se obţine cea mai mare simplificare a calculelor:

$$a = \frac{2\sum_{i=0}^{m} t_i^4 \sum_{i=-m}^{m} y_i - 2\sum_{i=0}^{m} t_i^2 \sum_{i=-m}^{m} t_i^2 y_i}{2n\sum_{i=0}^{m} t_i^4 - (2\sum_{i=0}^{m} t_i^2)^2} \qquad (23.23)$$

$$b = \frac{\sum_{i=-m}^{m} t_i y_i}{2\sum_{i=0}^{m} t_i^2} \qquad (23.24)$$

$$c = \frac{n\sum_{i=-m}^{m} t_i^2 y_i - 2\sum_{i=0}^{m} t_i^2 \sum_{i=-m}^{m} y_i}{2n\sum_{i=0}^{m} t_i^4 - (2\sum_{i=0}^{m} t_i^2)^2} \qquad (23.25)$$

23.2.6.3. **Tendinţa parabolică de gradul trei**

Determinarea parametrilor ecuaţiei de tendinţă, cu ajutorul metodei celor mai mici pătrate în cazul parabolei de gradul trei, presupune rezolvarea sistemului de ecuaţii normale:

$$\begin{aligned}
&na + b\sum_{i=-m}^{m} t_i + c\sum_{i=-m}^{m} t_i^2 + d\sum_{i=-m}^{m} t_i^3 = \sum_{i=-m}^{m} y_i \\
&a\sum_{i=-m}^{m} t_i + b\sum_{i=-m}^{m} t_i^2 + c\sum_{i=-m}^{m} t_i^3 + d\sum_{i=-m}^{m} t_i^4 = \sum_{i=-m}^{m} t_i y_i \\
&a\sum_{i=-m}^{m} t_i^2 + b\sum_{i=-m}^{m} t_i^3 + c\sum_{i=-m}^{m} t_i^4 + d\sum_{i=-m}^{m} t_i^5 = \sum_{i=-m}^{m} t_i^2 y_i \\
&a\sum_{i=-m}^{m} t_i^3 + b\sum_{i=-m}^{m} t_i^4 + c\sum_{i=-m}^{m} t_i^5 + d\sum_{i=-m}^{m} t_i^6 = \sum_{i=-m}^{m} t_i^3 y_i
\end{aligned} \qquad (23.26)$$

Deoarece sumele puterilor impare ale lui t_i sunt nule, sistemul (23.26) devine:

$$\begin{cases}
na + c\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} y_i \\
b\sum_{i=-m}^{m} t_i^2 + d\sum_{i=-m}^{m} t_i^4 = \sum_{i=-m}^{m} t_i y_i \\
a\sum_{i=-m}^{m} t_i^2 + c\sum_{i=-m}^{m} t_i^4 = \sum_{i=-m}^{m} t_i^2 y_i \\
b\sum_{i=-m}^{m} t_i^4 + d\sum_{i=-m}^{m} t_i^6 = \sum_{i=-m}^{m} t_i^3 y_i
\end{cases} \qquad (23.27)$$

Scanned with OKEN Scanner

Din prima şi a treia ecuaţie a sistemului (23.27) rezultă formulele de calcul simplificat ale parametrilor a şi c ai parabolei de gradul 3:

$$a = \frac{\sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^2 y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} \tag{23.28}$$

şi

$$c = \frac{n\sum_{i=-m}^{m} t_i^2 y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} y_i}{n\sum_{i=-m}^{m} t_i^4 - (\sum_{i=-m}^{m} t_i^2)^2} \tag{23.29}$$

iar din ecuaţia a doua şi a patra rezultă:

$$b = \frac{\sum_{i=-m}^{m} t_i^6 \sum_{i=-m}^{m} t_i y_i - \sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} t_i^3 y_i}{\sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^6 - (\sum_{i=-m}^{m} t_i^4)^2} \tag{23.30}$$

şi

$$d = \frac{\sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^3 y_i - \sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} t_i y_i}{\sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^6 - (\sum_{i=-m}^{m} t_i^4)^2} \tag{23.31}$$

Se observă că formulele parametrilor a şi c sunt identice cu cele obţinute anterior pentru parabola de gradul doi .

Exactitatea calculelor intermediare de ajustare se verifică cu ajutorul relaţiei:

$$\sum_{i=-m}^{m} y_i = \sum_{i=-m}^{m} y_{t_i}$$

Scanned with OKEN Scanner

În adevăr $$y_{t_i} = a + bt_i + ct_i^2$$

iar $$\sum_{i=-m}^{m} y_{t_i} = \sum_{i=-m}^{m} (a + bt_i + ct_i^2) = na + b\sum_{i=-m}^{m} t_i + c\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} y_i$$

În cazul general, tendinţa polinomială de gradul m depinde de $k = m+1$ parametri:

$$y_t = b_0 + b_1 t + b_2 t^2 + \cdots + b_m t^m \qquad (23.32)$$

Pentru determinarea celor k parametri, considerăm k puncte de coordonate $(y_1,t_1)\ldots(y_k,t_k)$ prin care presupunem că va trece curba y_t.

Punând condiţia ca polinomul (23.32) să treacă prin cele k puncte, se obţine sistemul de ecuaţii:

$$y_i = b_0 + b_1 t_i + \cdots + b_m t_i^m \qquad (i = 1,2,\ldots,k) \qquad (23.33)$$

Transcris sub formă matricială, sistemul (23.33) devine:

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_k \end{pmatrix}}_{Y} = \underbrace{\begin{pmatrix} 1 & t_1 & \cdots & t_1^m \\ 1 & t_2 & \cdots & t_1^m \\ 1 & t_3 & \cdots & t_3^m \\ \cdots & \cdots & \cdots & \cdots \\ 1 & t_k & \cdots & t_k^m \end{pmatrix}}_{T} \cdot \underbrace{\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{B} \qquad (23.34)$$

ceea ce ne conduce la matricea $T_{kx(m+1)}$, vectorul-coloană B_{mx1} al necunoscutelor şi vectorul-coloană Y_{mx1} al termenilor liberi.

Sintetic, relaţia (23.34) se poate scrie:

$$T_{kx(m+1)} \cdot B_{mx1} = Y_{mx1} \qquad (23.35)$$

de unde rezultă

Scanned with OKEN Scanner

$$B_{mx1} = T^{-1}_{kx(m+1)} \cdot Y_{mx1}$$

Notăm cu D determinantul matricei T:

$$D = \begin{vmatrix} 1 & t_1 & \dots t_1^m \\ \vdots & \vdots & \vdots \\ 1 & t_k & \dots t_k^m \end{vmatrix} \qquad (23.36)$$

şi cu $D_i (i = 0,1,2,\dots,m)$ determinanţii obţinuţi înlocuind coloana $i+1$ din matricea T cu vectorul-coloană Y:

$$D_i = \begin{vmatrix} 1 & t_1 \dots t_1^{i-1} & y_1 & t_1^{i+1} \dots t_1^m \\ \vdots & \vdots \quad \vdots & \vdots & \vdots \quad \vdots \\ 1 & t_k \dots t_k^{i-1} & y_k & t_k^{i+1} \dots t_k^m \end{vmatrix} \qquad (23.37)$$

Cu notaţiile (23.36) şi (23.37), formulele de calcul ale parametrilor $b_i \quad (i = 0,1,2,\dots,m)$ sunt, conform regulei lui *Cramer*:

$$b_i = \frac{D_1}{D} \qquad (i = 0,1,\dots,m) \qquad (23.38)$$

Ecuaţiile polinomiale aproximează bine evoluţia fenomenelor în timp pe perioade determinate. Totuşi, este greu să se găsească o bază logică pentru folosirea acestor ecuaţii de trend. Acest fapt este clar ilustrat de tendinţa liniară sau de cea descrisă de un polinom de orice grad care, prin prelungire poate deveni negativă într-o direcţie sau în ambele şi va creşte sau scădea nelimitat în ambele direcţii. Este clar că cele mai multe serii dinamice reale nu pot avea valori negative şi de regulă tind spre o valoare limită superioară sau inferioară, într-o direcţie sau în ambele.

Un polinom poate fi făcut să se apropie cât mai mult de evoluţia reală a unui fenomen prin creşterea gradului său. În particular, dacă polinomul are un grad egal cu numărul termenilor unei serii dinamice minus 1, atunci curba de tendinţă va trece prin toate punctele empirice, dar trendul va fi de neînţeles, deoarece numărul gradelor de libertate scade cu 1 pe măsură ce gradul polinomului creşte cu 1. Astfel, atunci când gradul polinomului este mai mic cu o unitate decât numărul termenilor unei serii dinamice, numărul gradelor de libertate ale polinomului este egal cu zero.

Scanned with OKEN Scanner

Dacă se adaugă noi termeni seriilor dinamice iniţiale, operaţia de ajustare trebuie reluată de la început.

23.2.6.4. Tendinţa exponenţială

Tendinţa exponenţială este specifică fenomenelor care se dezvoltă aproximativ în progresie geometrică şi este descrisă de funcţia:

$$y = ab^t$$

sau

$$lg\, y = lg\, a + t\, lg\, b \qquad (23.39)$$

Un fenomen se dezvoltă după modelul unei progresii geometrice în cazul în care indicii variaţiei cu bază mobilă sunt apropiaţi ca mărime între ei. Dacă $b > 1$ funcţia este crescătoare, iar dacă $0 < b < 1$ funcţia este descrescătoare.

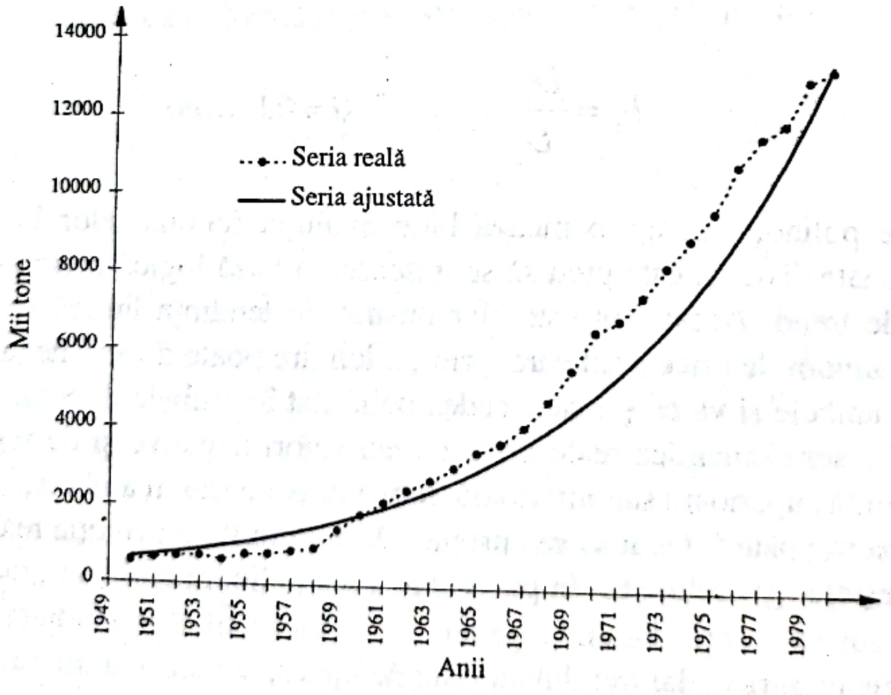


Figura 23.11. Producţia de oţel a *României* în perioada 1950 -1980

Scanned with OKEN Scanner

Tabelul 23.8. Ajustarea unei serii dinamice după o curbă exponenţială prin MCMMP

Anii	y_i	t_i	t_i^2	$\lg y_i$	$t_i \lg y_i$	$lgy_{t_i}=lga+tlgb$	y_{t_i}
1950	555	-15	225	2.74429	-41.16439	2.82131	662.7
1951	644	-14	196	2.80888	-39.32440	2.86472	732.4
1952	698	-13	169	2.84385	-36.97012	2.90813	809.3
1953	717	-12	144	2.85551	-34.26623	2.95154	894.4
1954	628	-11	121	2.79795	-30.77756	2.99495	988.4
1955	766	-10	100	2.88422	-38.84229	3.03835	1092.3
1956	779	-9	81	2.89153	-26.02384	3.08176	1207.2
1957	864	-8	64	2.93651	-23.49211	3.12517	1334.0
1958	934	-7	49	2.97034	-20.79243	3.16858	1474.3
1959	1420	-6	36	3.15228	-18.91373	3.21199	1629.2
1960	1806	-5	25	3.25671	-16.28359	3.25539	1800.5
1961	2126	-4	16	3.3275	-13.31025	3.29880	1989.8
1962	2451	-3	9	3.38934	-10.16803	3.34221	2198.9
1963	2707	-2	4	3.43248	-6.86498	3.38562	2430.1
1964	3039	-1	1	3.48273	-3.48273	3.42903	2685.5
1965	3426	0	0	3.53478	0.00000	3.47243	2967.8
1966	3670	1	1	3.56466	3.56467	3.51584	3279.8
1967	4088	2	4	3.61151	7.22302	3.55925	3624.5
1968	4751	3	9	3.67678	11.03036	3.60266	4005.5
1969	5540	4	16	3.74350	14.97404	3.64607	4426.6
1970	6517	5	25	3.81404	19.07024	3.68947	4891.9
1971	6803	6	36	3.83270	22.99620	3.73288	5406.1
1972	7401	7	49	3.86929	27.08503	3.77629	5974.3
1973	8161	8	64	3.91174	31.29395	3.81970	6602.3
1974	8848	9	81	3.94684	35.52161	3.86311	7296.4
1975	9549	10	100	3.97995	39.79958	3.90651	8063.3
1976	10733	11	121	4.03072	44.33793	3.94992	8910.9
1977	11457	12	144	4.05907	48.70885	3.99333	9847.6
1978	11779	13	169	4.07110	52.92441	4.03674	10882.7
1979	12909	14	196	4.11089	57.55250	4.08015	12026.7
1980	13175	15	225	4.11975	61.79626	4.12355	13290.9
Total	148941		2480	107.65165	127.20196	107.64545	133426.3

Scanned with OKEN Scanner

Aplicând metoda celor mai mici pătrate pentru estimarea parametrilor a şi b ai funcţiei (23.39), rezultă sistemul de ecuaţii normale:

$$\begin{cases} n\lg a + \lg b \sum_{i=0}^{n} t_i = \sum_{i=0}^{n} \lg y_i \\ \lg a \sum_{i=0}^{n} t_i + \lg b \sum_{i=0}^{n} t_i^2 = \sum_{i=0}^{n} t_i \lg y_i \end{cases} \tag{23.40}$$

Sistemul (23.40) se poate simplifica, dacă se modifică originea de calcul prin substituţia $\sum_{i=-m}^{m} t_i = 0$, astfel:

$$\begin{cases} n\lg a = \sum_{i=-m}^{m} \lg y_i \\ \lg b \sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} t_i \lg y_i \end{cases} \tag{23.41}$$

Din sistemul (23.41) rezultă formulele de calcul simplificat ale parametrilor ecuaţiei exponenţiale:

$$a = anti\lg\left(\frac{\sum_{i=-m}^{m} \lg y_i}{n}\right) \tag{23.42}$$

$$b = anti\lg\left(\frac{\sum_{i=-m}^{m} t_i \lg y_i}{\sum_{i=-m}^{m} t_i^2}\right) \tag{23.43}$$

Pentru verificarea exactităţii calculelor de ajustare se utilizează relaţia:

$$\sum_{i=-m}^{m} \lg y_i = \sum_{i=-m}^{m} \lg y_{t_i}$$

Scanned with OKEN Scanner

Prognoza dezvoltării fenomenelor sociale şi economice după modelul curbei exponenţiale trebuie să se facă cu multă prudenţă şi pentru orizonturi nu prea îndepărtate de timp. Chiar dacă curba exponenţială este potrivită pentru ajustarea unei serii dinamice, extrapolarea ei pentru orizonturi mari de timp poate duce la rezultate neverosimile.

23.2.7. Tendinţa exponenţială modificată

Tendinţele de dezvoltare a fenomenelor economice, care prezintă indici de variaţie cu bază mobilă în scădere şi tind asimptotic către o valoare limită superioară se pot determina cu ajutorul curbei exponenţiale modificate. Modificarea constă în adunarea algebrică a unei constante k termenilor iniţiali ai seriei dinamice, obţinându-se în acest fel o nouă serie dinamică apropiată mai mult de modelul unei progresii geometrice:

$$y = k + a \cdot b^t \qquad (23.44)$$

unde k este o asimptotă.

În funcţie de valorile parametrilor a şi b curba exponenţială modificată poate avea mai multe forme (fig. 23.12).

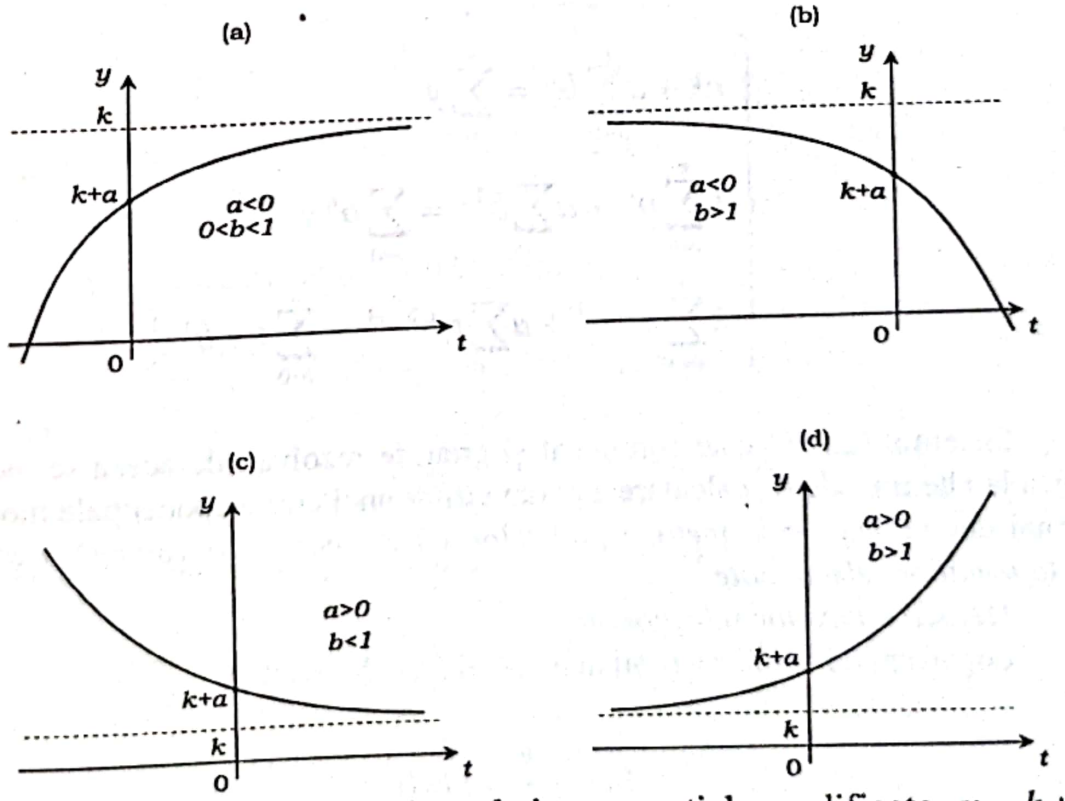


Figura 23.12. Variante ale curbei exponenţiale modificate $y = k + ab^t$

Scanned with OKEN Scanner

Varianta *(a)* descrie o curbă exponenţială modificată care prezintă o tendinţă de creştere cu o rată descrescătoare tinzând spre o valoare superioară limită k. Acesta este *singurul* model de trend exponenţial modificat care se aplică în practică. Varianta *(c)* se poate uneori utiliza ca model de trend în cazul seriilor dinamice descrescătoare cu o rată constantă de descreştere.

Prin transformări elementare funcţia (23.38) devine:

$$y - k = a \cdot b^t$$

iar

$$\frac{y-k}{a} = b^t$$

care se poate scrie

$$\frac{y-k}{a} = e^{t \ln b} \tag{23.45}$$

Dacă se aplică metoda celor mai mici pătrate pentru determinarea parametrilor curbei exponenţiale modificate, se obţine următorul sistem de ecuaţii normale:

$$\begin{cases} nk + a\sum_{i=0}^{n} b^{t_i} = \sum_{i=0}^{n} y_i \\ k\sum_{i=0}^{n} b^{t_i} + a\sum_{i=0}^{n} b^{2t_i} = \sum_{i=0}^{n} b^{t_i} y_i \\ k\sum_{i=0}^{n} t_i b^{t_i - 1} + a\sum_{i=0}^{n} t_i b^{2t_i - 1} = \sum_{i=0}^{n} t_i y_i b^{t_i - 1} \end{cases} \tag{23.46}$$

Sistemul (23.46) este complicat şi greu de rezolvat, de aceea se recurge în practică la alte metode de calculare a parametrilor unei curbe eponenţiale modificate. Cea mai des utilizată este *metoda punctelor selecţionate* şi o variantă a acesteia - *metoda mediilor selecţionate.*

Metoda punctelor selecţionate.

Logaritmând ambii membri ai ecuaţiei (23.45), rezultă:

$$\ln \frac{y-k}{a} = t \ln b \tag{23.47}$$

Scanned with OKEN Scanner

Pentru determinarea constantei k considerăm trei puncte echidistante, simetrice faţă de origine $(-t_1, y_1), (0, y_2), (t_1, y_3)$ şi, punând condiţia ca ecuaţia (23.47) să treacă prin aceste puncte, se obţine următorul sistem de ecuaţii:

$$\begin{cases} ln\dfrac{y_1 - k}{a} = -t_1 \, ln b \\ ln\dfrac{y_2 - k}{a} = 0 \\ ln\dfrac{y_3 - k}{a} = t_1 \, ln b \end{cases} \tag{23.48}$$

Din ecuaţia a doua a sistemului (23.48) rezultă:

$$\frac{y_2 - k}{a} = 1 \quad \text{sau} \quad y_2 - k = a \tag{23.49}$$

Din prima şi a treia ecuaţie a sistemului (23.48), prin însumare, se obţine:

$$ln\frac{y_1 - k}{a} + ln\frac{y_3 - k}{a} = 0 \tag{23.50}$$

adică

$$\frac{y_1 - k}{a} \cdot \frac{y_3 - k}{a} = 1$$

1

Efectuând calculele în (23.50) se obţine:

$$(y_1 - k) \cdot (y_3 - k) = a^2$$

sau

$$y_1 y_3 - k y_1 - k y_3 + k^2 = a^2 \tag{23.51}$$

Ţinând seama de relaţia (23.49), ecuaţia (23.51) devine:

$$y_1 y_3 - k y_1 - k y_3 + k^2 = y_2^2 - 2k y_2 + k^2$$

sau

$$y_1 y_3 - y_2^2 = k y_1 + k y_3 - 2k y_2$$

de unde

Scanned with OKEN Scanner

$$k = \frac{y_1 y_3 - y_2^2}{y_1 + y_3 - 2y_2} \qquad (23.52)$$

În formula (23.52) valorile y_1, y_2 şi y_3 pot fi termeni ai seriei, din centrul fiecărei treimi a ei.

Ajustarea unei serii după modelul curbei exponenţiale modificate prin metoda punctelor selecţionate presupune parcurgerea următoarelor etape de calcul:

1) Se împarte seria iniţială în trei părţi egale şi apoi se selecţionează din centrul fiecărei grupe de termeni valorile echidistante y_1, y_2 şi y_3. Dacă numărul termenilor seriei dinamice nu este divizibil cu trei se pot considera intervale care se suprapun uşor unul peste altul, cu condiţia ca distanţele dintre y_1, y_2 şi y_3 să fie egale între ele.

2) Se determină valoarea constantei k pe baza formulei (23.52).

3) Se determină termenii seriei empirice modificate prin adunarea algebrică a valorii k la termenii reali ai seriei date.

4) Se ajustează seria modificată după o funcţie exponenţială de forma :

$$y' = y - k = a \cdot b^t$$

5) Se recalculează valorile tendinţei prin scăderea algebrică a constantei k.

Pentru exemplificare să considerăm exemplul din tabelul (23.9) referitor la dinamica produsului social şi să ajustăm această serie dinamică după modelul unei curbe exponenţiale modificate. Deoarece numărul termenilor seriei nu este divizibil cu 3, renunţăm la primul termen. Cele trei puncte selecţionate vor avea coordonatele:

t_i	y_i
1951	125
1964	377
1977	1200

Substituind aceste date în formula (23.52) se obţine:

$$k = \frac{125 \cdot 1200 - 377^2}{125 + 1200 - 2 \cdot 377} = 13{,}785$$

Scanned with OKEN Scanner

Aceasta înseamnă că dinamica reală a produsului social se abate de la modelul unei progresii geometrice în plus cu $k = 13{,}785\%$. Scăzând constanta k din termenii reali ai seriei se obţine seria modificată (col. 3 a tabelului 23.9).

Tabelul 23.9. Ajustarea seriei dinamice după modelul curbei exponenţiale modificate

Anii	y_i %	$y_i'=y_i-k$	$\ln y_i$	t_i	t_i^2	$t_i \ln g y_i'$	$\ln y'_{t_i}$	y'_{t_i}	y_{t_i}
1	2	3	4	5	6	7	8	9	10
1951	125	111	2,0462	-13	169	-26,6001	2,0500	112	126
1952	137	123	2,0907	-12	144	-25,0880	2,0894	123	137
1953	157	143	2,1560	-11	121	-23,7159	2,1288	135	148
1954	161	147	2,1680	-10	100	-21,6795	2,1682	147	161
1955	186	172	2,2361	-9	81	-20,1246	2,2076	161	175
1956	187	173	2,2386	-8	64	-17,9087	2,2470	177	190
1957	207	193	2,2860	-7	49	-16,0023	2,2863	193	207
1958	215	201	2,3037	-6	36	-13,8220	2,3257	212	225
1959	237	23	2,3487	-5	25	-11,7436	2,3651	232	246
1960	263	249	2,3966	-4	16	-9,5863	2,4045	254	268
1961	293	279	2,4459	-3	9	-7,3378	2,4439	278	292
1962	311	297	2,4731	-2	4	-4,9461	2,4833	304	318
1963	338	324	2,5108	-1	1	-2,5108	2,5227	333	347
1964	377	363	2,5602	0	0	0,0000	2,5621	365	379
1965	414	400	2,6023	1	1	2,6023	2,6014	399	413
1966	457	443	2,6466	2	4	5,2932	2,6408	437	451
1967	501	487	2,6877	3	9	8,0632	2,6802	479	493
1968	540	526	2,7212	4	16	10,8847	2,7196	524	538
1969	582	568	2,7545	5	25	13,7726	2,7590	574	588
1970	629	615	2,7890	6	36	16,7342	2,7984	629	642
1971	701	687	2,8371	7	49	19,8596	2,8378	688	702
1972	771	757	2,8792	8	64	23,0338	2,8772	754	767
1973	852	838	2,9234	9	81	26,3102	2,9165	825	839
1974	947	933	2,9700	10	100	29,6998	2,9559	904	917
1975	1000	986	2,9940	11	121	32,9337	2,9953	989	1003
1976	1100	1086	3,0359	12	144	36,4310	3,0347	1083	1097
1977	1200	1186	3,0791	13	169	39,9641	3,0741	1186	1200
Total	-	-	69,1755	0	1638	64,5164	69,1755	-	-

După ce s-au determinat valorile transformate y_i', se ajustează seria astfel modificată după modelul curbei exponenţiale (23.44). Calculele intermediare se prezintă în tabelul (23.9).

Valorile parametrilor curbei exponenţiale modificate se determină tot pe baza formulelor (23.54) şi (23.55), astfel:

Scanned with OKEN Scanner

$$a = \text{antilg}\frac{69{,}1755}{27} = \text{antilg}\,2{,}56206 = 364{,}800$$

şi

$$b = \text{antilg}\frac{64{,}5164}{1638} = \text{antilg}\,0{,}03939 = 1{,}095$$

Deci, curba exponenţială modificată va fi:

$$y'_{t_i} = 364{,}8 \cdot 1{,}095^{t_i}$$

sau

$$lg\, y'_{t_i} = 2{,}56206 + 0{,}03939 t_i$$

Substituind succesiv în aceste formule valorile lui t_i, se vor obţine valorile seriei modificate ajustate din coloanele 8 şi 9 ale tabelului 23.9.

Considerând că şi în următorii ani dinamica produsului social va evolua după acelaşi model ca şi în perioada 1951-1977, s-a calculat prognoza până în anul 1985 (tabelul 23.10).

Tabelul 23.10. Prognoza dinamicii produsului social în perioada 1978-1985

Anii	Dinamica produsului social
1978	1298
1979	1421
1980	1556
1981	1704
1982	1866
1983	2043
1984	2237
1985	2450

Parametrii unei curbe exponenţiale modificate se pot determina şi cu ajutorul *metodei mediilor selecţionate*, mai simplu de calculat şi cu posibilităţi de generalizare şi pentru curba lui *Gompertz* şi curba logistică.

Pentru aceasta se împarte seria în trei părţi egale.

Considerăm o serie formată din $3n$ termeni şi împărţită în trei grupe de câte n termeni (tabelul 23.11).

Mediile celor trei perioade se determină astfel:

Scanned with OKEN Scanner

$$\bar{y}_1 = k + a \cdot \frac{\sum_{i=1}^{n} b^{t_i}}{n} \tag{23.53}$$

$$\bar{y}_2 = k + a \cdot \frac{\sum_{i=n+1}^{2n} b^{t_i}}{n} \tag{23.54}$$

$$\bar{y}_3 = k + a \cdot \frac{\sum_{i=2n+1}^{3n} b^{t_i}}{n} \tag{23.55}$$

Scăzând pe (23.53) din (23.54) şi pe (23.54) din (23.55) rezultă:

$$\bar{y}_2 - \bar{y}_1 = a \cdot \frac{\sum_{i=n+1}^{2n} b^{t_i} - \sum_{i=1}^{n} b^{t_i}}{n} \tag{23.56}$$

$$\bar{y}_3 - \bar{y}_2 = a \cdot \frac{\sum_{i=2n+1}^{3n} b^{t_i} - \sum_{i=n+1}^{2n} b^{t_i}}{n} \tag{23.57}$$

Raportând pe (23.57) la (23.56) obţinem:

$$\frac{\bar{y}_3 - \bar{y}_2}{\bar{y}_2 - \bar{y}_1} = \frac{\sum_{i=2n+1}^{3n} b^{t_i} - \sum_{i=n+1}^{2n} b^{t_i}}{\sum_{i=n+1}^{2n} b^{t_i} - \sum_{i=1}^{n} b^{t_i}} \tag{23.58}$$

de unde

$$b = \sqrt[n]{\frac{\bar{y}_3 - \bar{y}_2}{\bar{y}_2 - \bar{y}_1}} \tag{23.59}$$

Valoarea parametrului a rezultă din relaţia (23.56) sau din (23.57):

Scanned with OKEN Scanner

$$a = \frac{n \cdot (\bar{y}_2 - \bar{y}_1)}{\sum_{i=n+1}^{2n} b^{t_i} - \sum_{i=1}^{n} b^{t_i}} \quad \text{sau} \quad a = \frac{n \cdot (\bar{y}_3 - \bar{y}_2)}{\sum_{i=n+1}^{2n} b^{t_i} - \sum_{i=1}^{n} b^{t_i}} \tag{23.60}$$

Substituind acum pe a şi b în una dintre relaţiile (23.53), (23.54) sau (23.55), rezultă valoarea constantei k:

$$k = \bar{y}_1 - a \cdot \frac{\sum_{i=1}^{n} b^{t_i}}{n} \tag{23.61}$$

Tabelul 23.11. Metoda mediilor selecţionate

Anii (t_i)	y_i	$k + ab^t$	Mediile perioadelor
1 2 ⋮ n	y_1 y_2 ⋮ y_n	$k+ab^1$ $k+ab^2$ ⋮ $k+ab^n$	$\bar{y}_1$
Total	$\sum_{i=1}^{n} y_i$	$nk + a\sum_{i=1}^{n} b^{t_i}$	
$n+1$ $n+2$ ⋮ $2n$	y_{n+1} y_{n+2} ⋮ y_{2n}	$k+ab^{n+1}$ $k+ab^{n+2}$ ⋮ $k+ab^{2n}$	$\bar{y}_2$
Total	$\sum_{i=n+1}^{2n} y_i$	$nk + a\sum_{i=n+1}^{2n} b^{t_i}$	
$2n+1$ $2n+2$ ⋮ $3n$	y_{2n+1} y_{2n+2} ⋮ y_{3n}	$k+ab^{2n+1}$ $k+ab^{2n+2}$ ⋮ $k+ab^{3n}$	$\bar{y}_3$
Total	$\sum_{i=2n+1}^{3n} y_i$	$nk + a\sum_{i=2n+1}^{3n} b^{t_i}$	

Scanned with OKEN Scanner

Aplicarea practică a metodei mediilor selecţionate o vom efectua pe baza seriei din tabelul 23.9. Rezultatele intermediare şi cele finale sunt prezentate în tabelul 23.12.

Tabelul 23.12. Ajustarea seriei dinamice prin metoda mediilor selecţionate

Anii	Dinamica produsului social (%)	t_i	Mediile selecţionate	b^{t_i}	ab^{t_i}	$y_{t_i} = k + ab^{t_i}$
1	2	3	4	5	6	7
1951	125	0		1,000	109,987	125,951
1952	137	1		1,096	120,545	136,510
1953	157	2		1,201	132,094	148,058
1954	161	3		1,316	144,743	160,707
1955	186	4	$\bar{y}_1 = 179{,}111$	1,442	158,601	174,565
1956	187	5		1,581	173,889	189,853
1957	207	6		1,733	190,607	206,571
1958	215	7		1,899	208,865	224,829
1959	237	8		2,082	228,993	244,957
Total 1	1612	-		13,350		
1960	263	9		2,281	250,880	266,844
1961	293	10		2,500	274,968	290,932
1962	311	11		2,741	301,474	317,438
1963	338	12		3,004	330,401	346,365
1964	377	13	$\bar{y}_2 = 388{,}222$	3,292	362,077	378,041
1965	414	14		3,608	396,833	412,797
1966	457	15		3,954	434,888	450,853
1967	501	16		4,334	476,684	492,647
1968	540	17		4,750	522,438	538,402
Total 2	3494	-		30,464		
1969	582	18		5,206	572,592	588,556
1070	629	19		5,706	627,586	643,550
1971	701	20		6,254	687,859	703,823
1972	771	21		6,854	753,851	769,815
1973	852	22	$\bar{y}_3 = 864{,}666$	7,512	826,222	842,186
1974	947	23		8,233	905,523	921,487
1975	1000	24		9,023	992,413	1008,377
1976	1100	25		9,890	1087,771	1103,735
1977	1200	26		10,839	1192,149	1208,113
Total 3	7782	-		69,517		

Scanned with OKEN Scanner

$$b = \sqrt[9]{\frac{864,666 - 388,222}{388,222 - 179,111}} = \sqrt[9]{2,278}.$$

$$lg\, b = \frac{log\, 2,278}{9} = \frac{0,35755}{9} = 0,039733$$

de unde rezultă:

$$b = 1,096$$

$$a = \frac{9 \cdot (388,222 - 179,111)}{30,464 - 13,350} = \frac{1881,999}{17,111} = 109,987$$

şi

$$k = 179,111 - 109,987 \cdot \frac{13,350}{9} = 15,964$$

Seria iniţială şi cea ajustată se prezintă în figura 23.13.

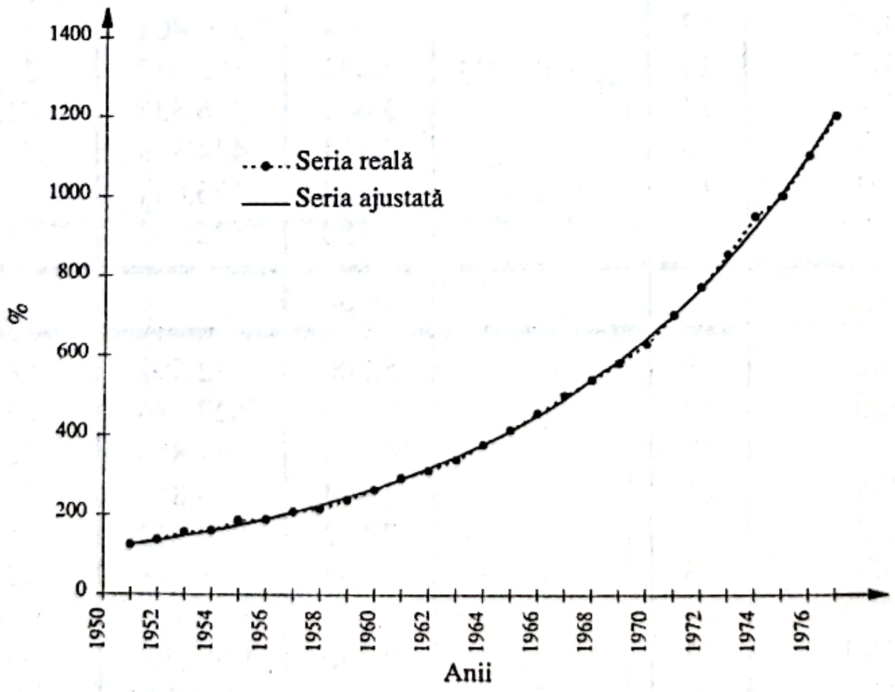


Figura 23.13. Dinamica produsului social în perioada 1951-1977

Scanned with OKEN Scanner

Înlocuind cu valorile lor pe a, b şi k în relaţia (23.44) rezultă:

$$y_{t_i} = 15{,}694 + 109{,}987 \cdot 1{,}096^{t_i}$$

Substituind în această ecuaţie valorile lui t_i din col. 3 a tabelului 23.12, rezultă valorile seriei ajustate din col. 7.

23.2.8. Curba lui Gompertz

Această curbă a fost introdusă în ştiinţă în anul 1825 de statisticianul englez *Benjamin Gompertz* (1779-1865) şi aplicată pe scară largă mai ales în demografie , pentru extrapolarea probabilităţilor de deces la vârstele înaintate; mai recent, ea se aplică în modelarea creşterii economice şi în marketing.

Curba lui *Gompertz* are forma unui *S alungit* şi descrie evoluţia unui fenomen ale cărui sporuri logaritmice de creştere se micşorează continuu cu o raţie constantă.

Formula obişnuită de calcul este:

$$y = a \cdot b^{c^t} \qquad \begin{cases} a < 0 \\ 0 < b < 1 \\ 0 \le c \le 1 \end{cases}$$

Curba lui *Gompertz* poate fi scrisă într-o formă similară curbei exponenţiale modificate, prin logaritmarea, astfel:

$$lg\, y = lg\, a + c^t\, lg\, b \qquad (23.62)$$

Primul termen din membrul al doilea al relaţiei (23.62), $\lg a$, reprezintă valoarea maximă sau *plafonul* către care tinde curba lui *Gompertz*, iar termenul al doilea reprezintă distanţa care separă, la un moment dat, nivelul atins de fenomenul luat în calcul de nivelul lui maxim. Desigur, această distanţă se micşorează din ce în ce mai mult pe măsură ce se înaintează în timp. (fig. 23.14).

Din relaţia (23.62) se observă că, dacă se face substituţia $\lg y = y'$, $\lg a = a'$ şi $\lg b = b'$, se obţine ecuaţia unei curbe exponenţiale modificate:

$$y' = a' + b'c^t \qquad a < c, \qquad c < b < 1 \qquad (23.63)$$

Scanned with OKEN Scanner

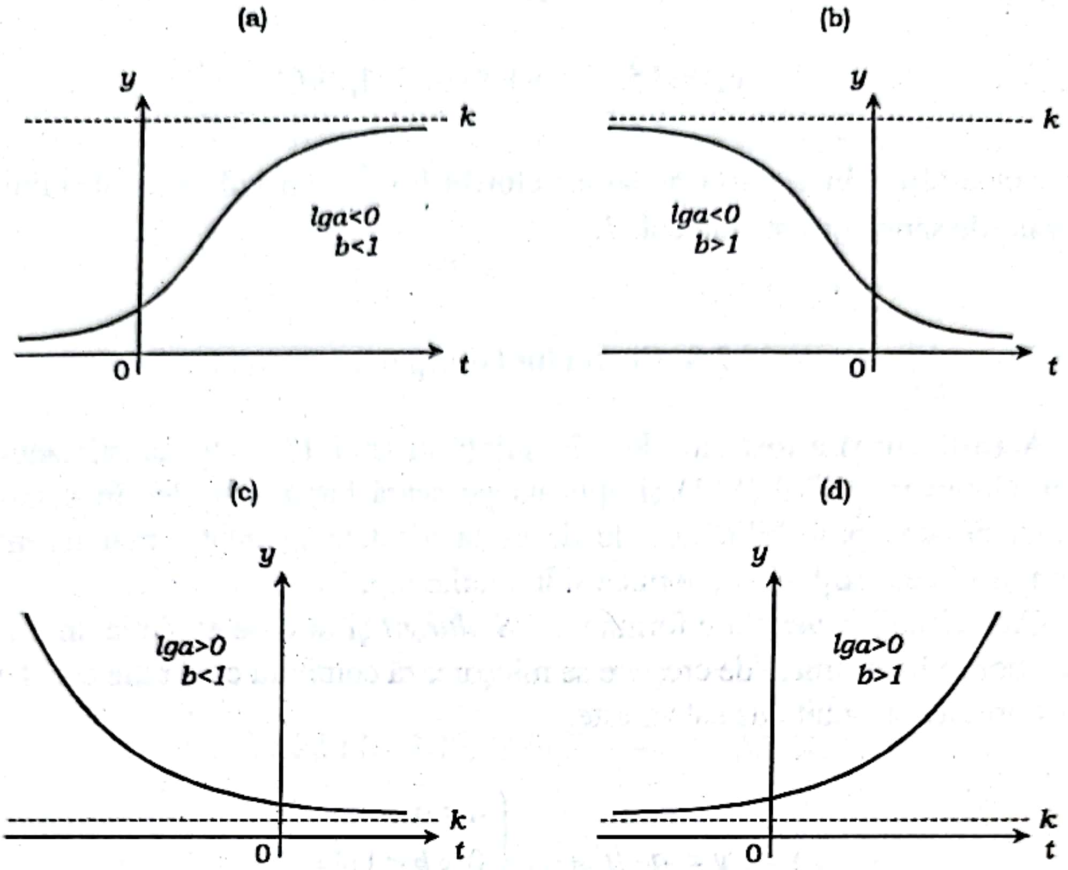


Figura 23.14. Forme ale unei curbe *Gompertz* $y = a \cdot b^{c^t}$

Curba lui *Gompertz* are proprietatea că primele diferențe ale valorilor $lg\, y_i$ au o rație constantă de creștere și egală cu b.

Dacă se reprezintă pe scară semilogaritmică ($\lg y_i$ pe ordonată și t_i pe abscisă), atunci funcția (23.63) ia formă de curbă concavă, ceea ce înseamnă că rata de creștere devine din ce în ce mai mică pe măsură ce t crește.

Reprezentată într-un sistem de axe rectangulare pe scară aritmetică, curba lui *Gompertz* ia forma unui *S* asimetric și are două asimptote: cea inferioară egală cu zero, iar cea superioară egală cu a'. Asimetria acestei curbe se explică prin comportamentul diferit al ei de o parte și de alta a punctului de inflexiune.

Curba lui *Gompertz* poate fi scrisă și sub formă liniară:

$$lg\, y = lg\, a + c^t\, lg\, b$$

Notând în (23.64) $lg\, y = y'$; $\lg a = a'$; $lg\, b = b'$ rezultă:

Scanned with OKEN Scanner

$$y' = a' + b'c^t$$

$$lg\frac{y}{a} = lg\,b' + t\,lg\,c$$

$$lg\,lg\frac{y'}{a'} = lg\,b' + t\,lg\,c$$

sau notând $A = \lg b'$ şi $B = lg\,c$ se obţine:.

$$lg\,lg\left(\frac{y}{a}\right) = A + Bt \tag{23.64}$$

Dacă în locul logaritmilor zecimali se vor folosi logaritmii naturali în ecuaţia (23.64), atunci nu se schimbă valorile lui y şi ale lui a, ci doar cele ale lui b şi c.

Ca şi în cazul curbei exponenţiale modificate, pentru determinarea parametrilor curbei Gompertz nu se poate utiliza metoda celor mai mici pătrate, care ar duce la un sistem de ecuaţii complicat şi deci greu de rezolvat. De aceea, se recomandă utilizarea în practică a metodei mediilor selecţionate.

În cazul curbei lui *Gompertz* aplicarea metodei mediilor selecţionate pentru determinarea parametrilor se poate face în mod asemănător cu cazul curbei exponenţiale modificate, dar numai după liniarizarea funcţiei pe baza formulei (23.62).

Parametrii curbei *Gompertz* vor avea următoarele expresii:

$$c = \sqrt[n]{\frac{\sum_{i=2n+1}^{3n} lg\,y_i - \sum_{i=n+1}^{2n} lg\,y_i}{\sum_{i=n+1}^{2n} lg\,y_i - \sum_{i=1}^{n} lg\,y_i}} \tag{23.65}$$

$$log\,b = \frac{\sum_{i=n+1}^{2n} lg\,y_i - \sum_{i=1}^{n} lg\,y_i}{\sum_{i=n+1}^{2n} c^{t_i} - \sum_{i=1}^{n} c^{t_i}} \tag{23.66}$$

Scanned with OKEN Scanner

$$\log a = \frac{\sum_{i=1}^{n} \lg y_i - \lg b \sum_{i=1}^{n} c^{t_i}}{n} \qquad (23.67)$$

Elementele de calcul necesare aplicării acestor formule se prezintă sistematizat în tabelul 23.13.

Tabelul 23.13. Aplicarea metodei mediilor selecţionate în cazul curbei lui *Gompertz*

Anii t_i	y_i	$\lg y_i$	Mediile selecţionate
1 2 ⋮ n	y_1 y_2 ⋮ y_n	$\lg y_1$ $\lg y_2$ ⋮ $\lg y_n$	$\frac{\sum_{i=1}^{n} \lg y_i}{n}$
Total 1	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} \lg y_i$	
$n+1$ $n+2$ ⋮ $2n$	y_{n+1} y_{n+2} ⋮ y_{2n}	$\lg y_{n+1}$ $\lg y_{n+2}$ ⋮ $\lg y_{2n}$	$\frac{\sum_{i=n+1}^{2n} \lg y_i}{n}$
Total 2	$\sum_{i=n+1}^{2n} y_i$	$\sum_{i=n+1}^{2n} \lg y_i$	
$2n+1$ $2n+2$ ⋮ $3n$	y_{2n+1} y_{2n+2} ⋮ y_{3n}	$\lg y_{2n+1}$ $\lg y_{2n+2}$ ⋮ $\lg y_{3n}$	$\frac{\sum_{i=2n+1}^{3n} \lg y_i}{n}$
Total 3	$\sum_{i=2n+1}^{3n} y_i$	$\sum_{i=2n+1}^{3n} \lg y_i$	

Scanned with OKEN Scanner

Se consideră, pentru exemplificarea calculului parametrilor curbei lui *Gompertz*, seria dinamică din tabelul 23.14, cu privire la evoluţia producţiei de oţel a *României* în perioada 1950-1976. Reprezentând grafic seria (fig. 23.15) se observă că ea se apropie de forma unei curbe *Gompertz*, depăşind puţin punctul ei de inflexiune. Calculele preliminare necesare pentru determinarea parametrilor a, b şi c se prezintă sistematizat în tabelul 23.14.

Tabelul 23.14. Determinarea curbei *Gompertz* de evoluţie a producţiei de oţel a *României* în perioada 1950-1976

Anii	Mii tone y_i	$\lg y_i$	t_i	c^{t_i}	$c^{t_i} \cdot \lg b$	$\lg y_{t_i}$	y_{t_i}
1	2	3	4	5	6	7	8
1950	555	2,74429	0	1,00000	-3,18406	2,57654	377
1951	644	2,80899	1	0,97657	-3,10945	2,65116	448
1952	698	2,84385	2	0,95368	-3,03658	3,72402	530
1953	717	2,85552	3	0,93133	-2,96542	2,79519	624
1954	628	2,79796	4	0,90950	-2,89592	2,86468	732
1955	766	2,88423	5	0,88819	-2,82806	2,93255	856
1956	779	2,89154	6	0,86732	-2,76178	2,99882	997
1957	864	2,93651	7	0,84705	-2,69705	3,06355	1158
1958	934	2,97035	8	0,82720	-2,63385	3,12675	1339
1959	1420	3,15229	9	0,80781	-2,57213	3,18847	1543
1960	1806	3,25672	10	0,78888	-2,51185	3,24875	1773
1961	2126	3,32756	11	0,77039	-2,45298	3,30762	2031
1962	2451	3,38934	12	0,75234	-2,39550	3,36510	2318
1963	2704	3,43201	13	0,73471	-2,33936	3,42124	2638
1964	3039	3,48273	14	0,71749	-2,28454	3,47606	2993
1965	3426	3,53479	15	0,70068	-2,23100	3,52960	3385
1966	3670	3,56467	16	0,68426	-2,17872	3,58189	3818
1967	4088	3,61151	17	0,66822	-2,12766	3,63294	4295
1968	4751	3,67678	18	0,65256	-2,07780	3,68281	4817
1969	5540	3,74351	19	0,63727	-2,02910	3,73150	5389
1970	6517	3,81405	20	0,62233	-1,98155	3,77905	6012
1971	6803	3,83270	21	0,60775	-1,93511	3,82549	6691
1972	7401	3,86929	22	0,59351	-1,88977	3,87084	7427
1973	8161	3,91174	23	0,57960	-1,84548	3,91512	8225
1974	8848	3,94685	24	0,56602	-1,80223	3,95837	9086
1975	9549	3,97996	25	0,55275	-1,76000	4,00061	10014
1976	10733	4,03072	26	0,53980	-1,71875	4,04185	11012

Sursa: Anuarele statistice ale României

Scanned with OKEN Scanner

S-a împărţit seria în trei părţi egale formate din câte 9 termeni şi apoi s-au determinat sumele logaritmilor valorilor lui y pentru fiecare parte, obţinându-se:

$$\sum_{i=0}^{n-1} lg\, y_i = 25{,}73314$$

$$\sum_{i=n}^{2n-1} lg\, y_i = 30{,}75162$$

$$\sum_{i=2n}^{3n-1} lg\, y_i = 34{,}80560$$

Substituind aceste valori în formulele (23.65), (23.66) şi (23.67) rezultă:

$$c = 0{,}97657$$

$$lg\, b = -3{,}18406$$

$$lg\, a = 5{,}76060$$

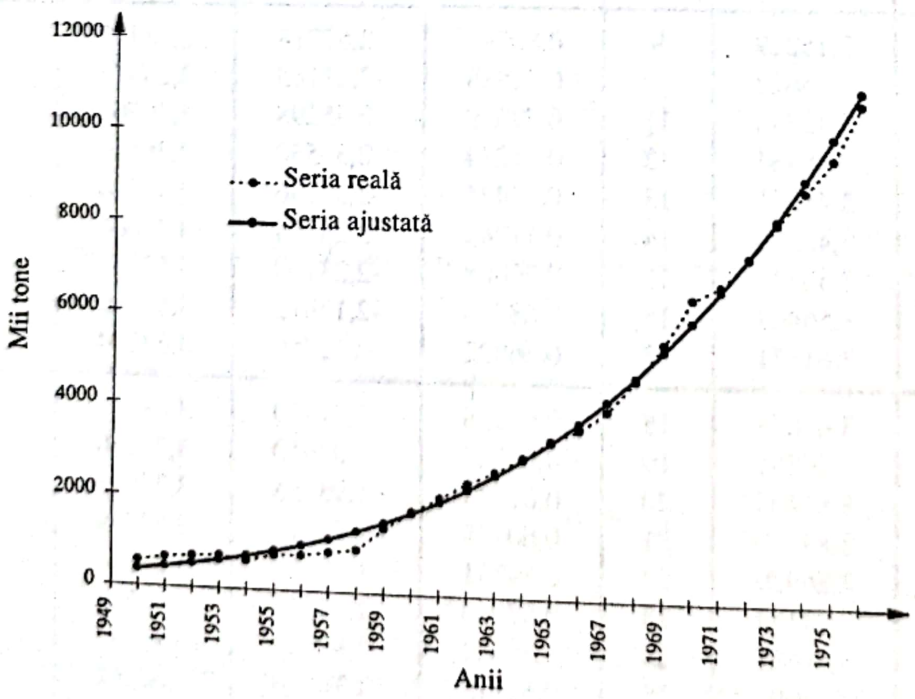


Figura 23.15. Evoluţia producţiei de oţel a *României* în perioada 1950-1976

Scanned with OKEN Scanner

Aceasta înseamnă că evoluţia producţiei de oţel a *României* în perioada 1950 -1976 poate fi descrisă de următoarea ecuaţie *Gompertz:*

$$lg\, y_t = 5{,}76060 + 0{,}97657^{t_i} \cdot (-3{,}18406)$$

Substituind succesiv valorile lui t_i în ecuaţia de mai sus, rezultă valorile ajustate ale producţiei de oţel, astfel:

$$y_{1950} = \text{antilg}\,(5{,}76060 - 1 \cdot 3{,}18406) = 377 \text{ mii tone}$$
$$y_{1951} = \text{antilg}\,(5{,}76060 - 0{,}97657 \cdot 3{,}18406) = 448 \text{ mii tone}$$
$$\ldots\ldots\ldots

doua printr-un ritm descrescător care în final tinde spre zero. Dezvoltarea unui fenomen după o curbă logistică are loc între două limite: una inferioară, care se confundă cu axa absciselor (*minimul absolut* după *Verhulst*) şi una superioară (*maximul absolut*) care se stabileşte prin ajustare.

Ecuaţia curbei logistice este:

$$y = \frac{b}{1 + ae^{-Bt}} \tag{23.68}$$

Curba logistică dată de ecuaţia (23.68) are ca asimptotă inferioară dreapta $y = 0$, iar ca asimptotă superioară $y = b$ şi punctul de inflexiune $t = 0$ şi $y = \frac{b}{2}$ (fig. 23.16) şi (23.17).

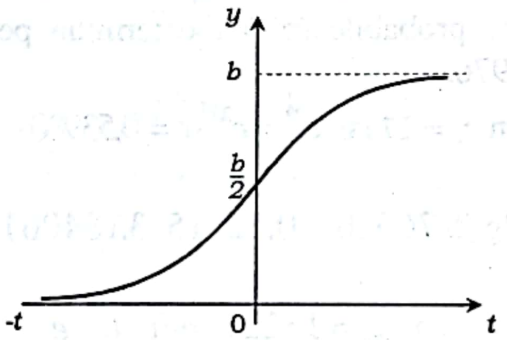


Figura 23.16. Curba logistică $y = \frac{b}{1 + ae^{-Bt}}$

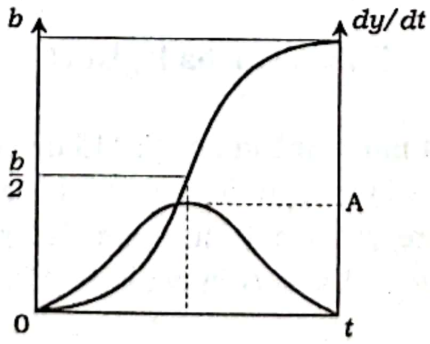


Figura 23.17. Curba logistică şi ritmul de creştere a unei variabile logistice

Scanned with OKEN Scanner

Prin transformări elementare, relaţia (23.68) devine:

$$ln\frac{b-y}{y} = -Bt + C \quad \text{unde} \quad C = \ln a \tag{23.69}$$

Pentru determinarea parametrilor B, b şi C se poate utiliza metoda punctelor selecţionate. Selectând trei puncte echidistante în timp $(0, y_1), (t, y_2), (2t, y_3)$ şi punând curbei (23.69) condiţia de a trece prin cele trei puncte rezultă sistemul de ecuaţii:

$$\begin{cases} ln\dfrac{b-y_1}{y_1} = C \\ ln\dfrac{b-y_2}{y_2} = -Bt + C \\ ln\dfrac{b-y_3}{y_3} = -2Bt + C \end{cases} \tag{23.70}$$

Din prima ecuaţie a sistemului (23.70) rezultă valoarea parametrului C

$$C = ln\frac{b-y_1}{y_1} = \frac{1}{0{,}43429} lg\frac{b-y_1}{y_1} \tag{23.71}$$

Înmulţind cu -2 a doua ecuaţie a sistemului (23.70) şi apoi adunând cele trei ecuaţii, se elimină B şi C şi rezultă:

$$ln\frac{b-y_1}{y_1} - 2\,ln\frac{b-y_2}{y_2} + ln\frac{b-y_3}{y_3} = 0$$

$$\frac{b-y_1}{y_1} \cdot \frac{b-y_3}{y_3} = \left(\frac{b-y_2}{y_2}\right)^2$$

efectuând apoi calculele se obţine:

$$b^2(y_2^2 - y_1y_3) - b(y_1y_2^2 + y_3y_2^2 - 2y_1y_2y_3) = 0$$

Deoarece soluţia $b = 0$ este exclusă, rezultă:

Scanned with OKEN Scanner

$$b = \frac{2y_1y_2y_3 - y_2^2(y_1 + y_3)}{y_1y_3 - y_2^2} \tag{23.72}$$

Valoarea parametrului C se obţine substituind pe b în formula (23.71). În sfârşit, scăzând primele două ecuaţii ale sistemului (23.70) obţinem valoarea parametrului B:

$$ln\frac{b-y_1}{y_1} - ln\frac{b-y_2}{y_2} = Bt$$

sau

$$ln\frac{b-y_1}{y_1} \cdot \frac{y_2}{b-y_2} = Bt$$

adică

$$B = \frac{1}{0{,}43429} lg\frac{(b-y_1)y_2}{y_1(b-y_2)} \tag{23.73}$$

Valorile celor trei parametri ai curbei logistice, determinate pe baza metodei punctelor selecţionate, depind numai de coordonatele celor trei puncte selecţionate şi fac abstracţie de toate celelalte puncte ale evoluţiei reale a fenomenului studiat. Acest fapt constituie principala deficienţă a metodei punctelor selecţionate.

O altă formă a curbei logistice, utilizată pentru determinarea tendinţelor generale de dezvoltare a fenomenelor economice, este dată de următoarea ecuaţie:

$$\frac{1}{y_{t_i}} = a + bc^{t_i} \tag{23.74}$$

Din această relaţie se observă asemănarea ecuaţiei curbei logistice cu cea a unei curbe exponenţiale modificate (23.44) cu deosebirea că, pentru variabila dependentă în cazul unei curbe logistice se consideră în locul valorilor ei reale, mărimile inverse ale acestora, sporurile cu bază mobilă ale lui $\frac{1}{y}$ scad cu o rată constantă.

Parametrii a, b şi c ai curbei logistice (23.74) diferă de cei ai formei clasice (23.68), dar ambele ecuaţii descriu curbe de forme asemănătoare şi se pot determina pe baza metodei mediilor selecţionate cu ajutorul formulelor:

Scanned with OKEN Scanner

$$c = \sqrt[n]{\frac{\sum_{i=2n+1}^{3n} \frac{1}{y_i} - \sum_{i=n+1}^{2n} \frac{1}{y_i}}{\sum_{i=n+1}^{2n} \frac{1}{y_i} - \sum_{i=1}^{n} \frac{1}{y_i}}} \qquad (23.75)$$

$$b = \frac{\sum_{i=n+1}^{2n} \frac{1}{y_i} - \sum_{i=1}^{n} \frac{1}{y_i}}{\sum_{i=n+1}^{2n} c^{t_i} - \sum_{i=1}^{n} c^{t_i}} \qquad (23.76)$$

$$a = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{y_i} - \frac{b}{n}\sum_{i=1}^{n} c^{t_i} \qquad (23.77)$$

Sistematizat, calculele de ajustare a unei serii dinamice după metoda mediilor selecţionate se prezintă în tabelul 23.15.

Întrucât seria este formată din 27 de termeni (număr divizibil prin 3), o vom împărţi în trei părţi egale, formate din câte $n = 9$ termeni, respectiv corespunzătoare perioadelor 1950-1958, 1959-1967 şi 1968-1976.

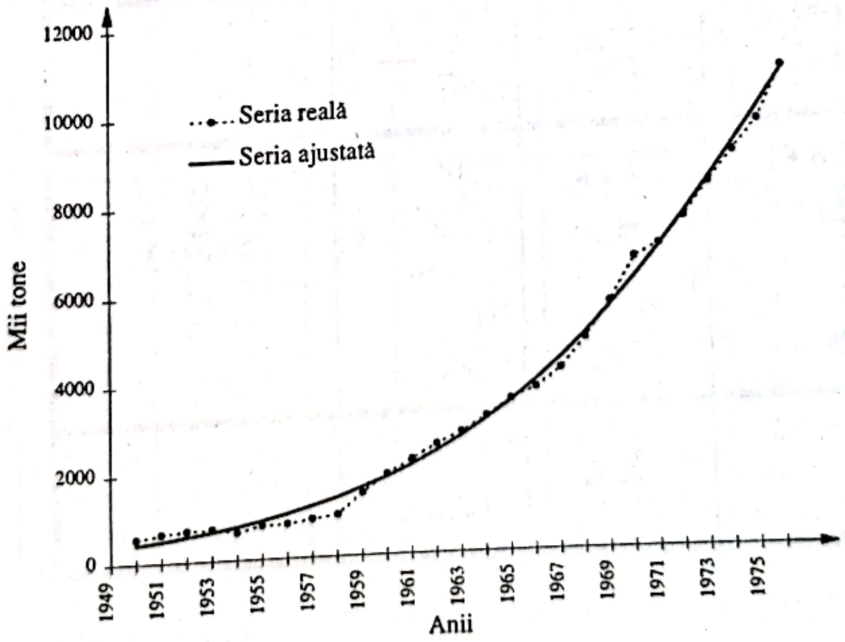


Figura 23.18. Producţia de oţel a *României* în perioada 1950-1976

Scanned with OKEN Scanner

Tabelul 23.15. Ajustarea seriilor dinamice (23.74) după metoda mediilor selecţionate

Anii t_i	y_i	$\frac{1}{y_i}$	Mediile selecţionate
1	2	3	4
1 2 $\vdots$ n	y_1 y_2 $\vdots$ y_n	$1/y_1$ $1/y_2$ $\vdots$ $1/y_n$	$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i}$
Total 1	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n}\frac{1}{y_i}$	
$n+1$ $n+2$ $\vdots$ $2n$	y_{n+1} y_{n+2} $\vdots$ y_{2n}	$1/y_{n+1}$ $1/y_{n+2}$ $\vdots$ $1/y_{2n}$	$\frac{1}{n}\sum_{i=n+1}^{2n}\frac{1}{y_i}$
Total 2	$\sum_{i=n+1}^{2n} y_i$	$\sum_{i=n+1}^{2n}\frac{1}{y_i}$	
$2n+1$ $2n+2$ $\vdots$ $3n$	y_{2n+1} y_{2n+2} $\vdots$ y_{3n}	$1/y_{2n+1}$ $1/y_{2n+2}$ $\vdots$ $1/y_{3n}$	$\frac{1}{n}\sum_{i=2n+1}^{3n}\frac{1}{y_i}$
Total 3	$\sum_{i=2n+1}^{3n} y_i$	$\sum_{i=2n+1}^{3n}\frac{1}{y_i}$	

Scanned with OKEN Scanner

Tabelul 23.16. Determinarea curbei logistice de evoluţie a producţiei de oţel a României în perioada 1950-1976

Anii	y_i	$\frac{1000}{y_i}$	t_i	c^{t_i}	bc^{t_i}	$a+bc^{t_i}$	y_{t_i}	$y_i - y_{t_i}$
1	2	3	4	5	6	7	8	9
1950	555	1,8018	0	1,0000	2,2725	2,3162	432	123
1951	644	1,5528	1	0,8633	1,9619	2,0057	499	145
1952	698	1,4327	2	0,7454	1,6938	1,7376	576	122
1953	717	1,3947	3	0,6435	1,4623	1,5061	664	53
1954	628	1,5924	4	0,5556	1,2625	1,3062	766	-138
1955	766	1,3055	5	0,4796	1,0899	1,1337	882	-116
1956	779	1,2837	6	0,4141	0,9410	0,9848	1015	-236
1957	864	1,1574	7	0,3575	0,8124	0,8562	1168	-304
1958	934	1,0707	8	0,3086	0,7014	0,7451	1342	-408
1959	1420	0,7042	9	0,2665	0,6055	0,6493	1540	-120
1960	1806	0,5537	10	0,2300	0,5224	0,5665	1765	41
1961	2126	0,4704	11	0,1986	0,4513	0,4951	2020	106
1962	2451	0,4080	12	0,1715	0,3897	0,4334	2307	144
1963	2704	0,3694	13	0,1480	0,3364	0,3802	2630	74
1964	3039	0,3291	14	0,1278	0,2904	0,3342	2992	47
1965	3426	0,2919	15	0,1103	0,2507	0,2945	3395	31
1966	3670	0,2725	16	0,0953	0,2165	0,2602	3843	-173
1967	4088	0,2446	17	0,0822	0,1869	0,2307	4335	-247
1968	4751	0,2105	18	0,0710	0,1614	0,2051	4875	-124
1969	5540	0,1805	19	0,0613	0,1393	0,1831	5462	78
1970	6517	0,1534	20	0,0529	0,1203	0,1640	6096	421
1971	6803	0,1470	21	0,0457	0,1038	0,1476	6775	28
1972	7401	0,1351	22	0,0394	0,0896	0,1334	7496	-95
1973	8161	0,1225	23	0,0341	0,0774	0,1212	8254	-93
1974	8848	0,1130	24	0,0294	0,0668	0,1106	9043	-195
1975	9549	0,1047	25	0,0254	0,0577	0,1015	9857	-308
1976	10733	0,0932	26	0,0219	0,0498	0,0936	10687	46

Sumele celor trei grupe de termeni înmulţiţi cu 1000 sunt:

Scanned with OKEN Scanner

$$\sum_{i=0}^{n-1}\frac{1000}{y_i}=12,592$$

$$\sum_{i=n}^{2n-1}\frac{1000}{y_i}=3,644$$

$$\sum_{i=2n}^{3n-1}\frac{1000}{y_i}=1,260$$

Substituind aceste valori în formulele (23.75), (23.76) şi (23.77), rezultă:

$$c=\sqrt[9]{\frac{1,260-3,644}{3,644-12,592}}=\sqrt[9]{0,0266}=0,863$$

$$b=\frac{3,644-12,592}{1,4302-5,3476}=2,272$$

$$a=\frac{1}{9}\cdot 12,592-\frac{2,272}{9}\cdot 5,3476=0,0044$$

Substituind valorile celor trei parametri în ecuaţia (23.74), rezultă valorile teoretice ale variabilei dependente, care se prezintă în col. 8 a tabelului 23.16

Ecuaţia curbei logistice pentru creşterea producţiei de oţel a *României* în perioada 1950-1976 are forma:

$$\frac{1}{y_t}=0,044+2,272\cdot 0,863^{t_i}$$

$t_0=0;\quad y_{1950}=1/(0,044+2,272\cdot 0,863^0)=432$ mii tone

$t_1=1;\quad y_{1951}=1/(0,044+2,272\cdot 0,863^1)=499$ mii tone

.....

$t_{26}=26;\quad y_{1976}=1/(0,044+2,272\cdot 0,863^{26})=10687$ mii tone

În ipoteza că şi în viitorul apropiat (după anul 1976) producţia de oţel va fi determinată de influenţa aceloraşi factori care au acţionat în perioada 1950-1976, prognoza producţiei de oţel se poate face prin extrapolare, astfel:

Scanned with OKEN Scanner

$t_{27} = 27;$ $y_{1977} = 1/(0{,}044 + 2{,}272 \cdot 0{,}863^{27}) = 11500$ mii tone

$t_{28} = 28;$ $y_{1978} = 1/(0{,}044 + 2{,}272 \cdot 0{,}863^{28}) = 12333$ mii tone

$t_{29} = 29;$ $y_{1979} = 1/(0{,}044 + 2{,}272 \cdot 0{,}863^{29}) = 13156$ mii tone

$t_{30} = 30;$ $y_{1980} = 1/(0{,}044 + 2{,}272 \cdot 0{,}863^{30}) = 13960$ mii tone

23.2.10. Curba exponenţială de gradul doi

Curba exponenţială de gradul doi este descrisă de ecuaţia:

$$y = A \cdot B^t \cdot C^{t^2} \tag{23.78}$$

Logaritmând ecuaţia (23.78) rezultă:

$$lg\, y = lg\, A + t\, lg\, B + t^2\, lg\, C \tag{23.79}$$

Dacă notăm: $a = lg\, A, b = lg\, B$ şi $c = lg\, C$, ecuaţia (23.79) devine:

$$lg\, y = a + bt + ct^2 \tag{23.80}$$

care se mai numeşte şi *parabolă logaritmică.*

Determinarea parametrilor ecuaţiei exponenţiale de gradul doi (23.79) este similară cu determinarea parametrilor parabolei de gradul doi, cu excepţia faptului că se lucrează cu logaritmii valorilor lui y în locul valorilor reale ale acestei variabile.

Aplicând metoda celor mai mici pătrate rezultă următorul sistem de ecuaţii normale:

$$\begin{cases} na + b\sum_{i=0}^{n} t_i + c\sum_{i=0}^{n} t_i^2 = \sum_{i=0}^{n} lg\, y_i \\ a\sum_{i=0}^{n} t_i + b\sum_{i=0}^{n} t_i^2 + c\sum_{i=0}^{n} t_i^3 = \sum_{i=0}^{n} t_i\, lg\, y_i \\ a\sum_{i=0}^{n} t_i^2 + b\sum_{i=0}^{n} t_i^3 + c\sum_{i=0}^{n} t_i^4 = \sum_{i=0}^{n} t_i^2\, lg\, y_i \end{cases} \tag{23.81}$$

Pentru a se simplifica calculul parametrilor a, b şi c se schimbă originea de calcul prin substituţia sumelor impare ale lui t_i cu zero în (23.81):

Scanned with OKEN Scanner

$$\begin{cases} na + c\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} \lg y_i \\ b\sum_{i=-m}^{m} t_i^2 = \sum_{i=-m}^{m} t_i \lg y_i \\ a\sum_{i=-m}^{m} t_i^2 + c\sum_{i=-m}^{m} t_i^4 = \sum_{i=-m}^{m} t_i^2 \lg y_i \end{cases} \quad (23.82)$$

Din ecuaţia a doua a sistemului (23.82) obţinem:

$$b = \frac{\sum_{i=-m}^{m} t_i \lg y_i}{\sum_{i=-m}^{m} t_i^2} \quad (23.83)$$

iar din prima parte şi a treia ecuaţie a aceluiaşi sistem, rezultă:

$$a = \frac{\sum_{i=-m}^{m} t_i^4 \sum_{i=-m}^{m} lg\, y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} t_i^2 lg\, y_i}{n\sum_{i=-m}^{m} t_i^4 - \left(\sum_{i=-m}^{m} t_i^2\right)^2} \quad (23.84)$$

$$c = \frac{n\sum_{i=-m}^{m} t_i^2 lg\, y_i - \sum_{i=-m}^{m} t_i^2 \sum_{i=-m}^{m} lg\, y_i}{n\sum_{i=-m}^{m} t_i^4 - \left(\sum_{i=-m}^{m} t_i^2\right)^2} \quad (23.85)$$

Determinarea tendinţei centrale de dezvoltare a fenomenelor este importantă atât pentru descrierea evoluţiei de lungă durată a acestora, cât şi pentru estimarea influenţei factorilor care determină o anumită evoluţie (cum ar fi factorii cu acţiune sezonieră ciclică şi factorii întâmplători).

Ajustând o serie dinamică nu trebuie să pierdem din vedere faptul că trebuie să luăm în calcul o perioadă suficient de lungă din trecut. Tendinţa centrală determinată pe baza unui număr redus de termeni este lipsită de semnificaţie şi neindicată pentru prognoză (extrapolare). Luând în calcul un număr mai mare de termeni este posibil să se determine adevăratele cauze ale evoluţiei, altele decât variabila independentă timp.

Scanned with OKEN Scanner

Experienţa statistică a arătat că *orizontul prognozei* obţinute pe baza tendinţei trebuie să fie foarte mic în comparaţie cu durata luată în calculul tendinţei. Astfel, de exemplu, dacă tendinţa s-a determinat pentru o perioadă de 50 ani, atunci orizontul prognozei trebuie să fie mai mic de un deceniu.

Calitatea ajustării se poate aprecia pe baza comparaţiei între abaterea medie pătratică a valorilor teoretice y_t faţă de valorile empirice y_i şi abaterea medie pătratică a valorilor empirice faţă de valoarea medie $\bar{y}$. Dacă perioada de timp luată în calcul este suficient de mare, atunci σ_{y_t} este mult mai mică decât σ_y.

23.3. Interpolarea seriilor dinamice

Interpolarea în cazul unei serii dinamice înseamnă determinarea valorilor indicatorului y pentru valori intermediare ale timpului t.

Fie o serie dinamică:

$$Y: \begin{pmatrix} t_0, t_1, t_2, \ldots, t_i, \ldots, t_n \\ y_0, y_1, y_2, \ldots, y_i, \ldots, y_n \end{pmatrix}$$

Teoretic, prin punctele de coordonate (t_i, y_i) pot trece un număr infinit de funcţii, în practică, însă, se poate considera că funcţia necunoscută îmbracă forma unui polinom de grad n, bine determinat, care pentru cele $n+1$ valori ale timpului va lua $n+1$ valori ale lui y:

$$P(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n \qquad (23.86)$$

Pe baza polinomului $P(t)$ se pot determina valorile lui pentru oricare valoare intermediară a lui t. O astfel de determinare a valorilor lui y_i, pentru orice valoare intermediară y_i poartă denumirea de *interpolare polinomială* sau *parabolică*.

Scanned with OKEN Scanner

23.3.1. Metoda lui Lagrange

Formula cea mai generală de interpolare pe baza unui polinom este cea a lui *Lagrange*[1]:

$$P(t) = \sum_{k=0}^{n} \frac{\prod^{(k)} (t - t_i)}{\prod^{(k)} (t_k - t_i)} \cdot y_k \qquad (k \neq i) \tag{23.87}$$

Dacă notăm $F = (t - t_0)(t - t_1)\ldots(t - t_n)$ formula lui *Lagrange* devine:

$$P(t) = \sum_{i=0}^{n} \frac{F(t)}{F(t_k)} \cdot \frac{y_k}{t - t_k} \tag{23.88}$$

Dacă în formula (24.2.) considerăm $k = 0$ şi $n = 1$ rezultă:

$$P(t) = \frac{t - t_1}{t_0 - t_1} \cdot y_0 + \frac{t - t_0}{t_1 - t_0} \cdot y_1 \tag{23.89}$$

sau

$$P(t) = y_0 + \frac{y_0 - y_1}{t_0 - t_1} \cdot (t - t_0) \tag{23.90}$$

Formula (24.4) se aplică în practică atunci când cunoaştem două valori ale timpului şi valorile corespunzătoare ale lui y şi poartă denumirea de formula de *interpolare liniară.*

Din punct de vedere geometric, interpolarea liniară înseamnă înlocuirea curbei care trece prin cele două puncte de coordonate cunoscute cu coarda care le uneşte. Un punct intermediar de pe abscisă t $(t_0 < t < t_1)$ va avea ordonata $P(t)$ dată de formula (24.4).

[1] Joseph Louis Lagrange (1736 - 1813) - matematician şi fizician francez

Scanned with OKEN Scanner

23.3.2. Metoda lui *Newton*

Fie o serie dinamică:

$$Y: \begin{pmatrix} t_0, t_1, t_2, \ldots, t_i, \ldots, t_n \\ y_0, y_1, y_2, \ldots, y_i, \ldots, y_n \end{pmatrix}$$

în care t_i variază în progresie aritmetică cu raţia r :

$$t_0, t_0 + r, t_0 + 2r, \ldots, t_0 + nr$$

atunci polinomul de gradul n care trece prin cele $n+1$ puncte va avea expresia:

$$P(t) = y_0 + \frac{t - t_0}{r} \cdot \frac{y_0}{1!} + \left(\frac{t - t_0}{r}\right)\left(\frac{t - t_0}{r} - 1\right) \cdot \frac{\Delta^2_{y_0}}{2!} + \cdots +$$

$$+ \left(\frac{t - t_0}{r}\right)\left(\frac{t - t_0}{r} - 1\right) \cdots \left(\frac{t - t_0}{r} - n + 1\right) \cdot \frac{\Delta^n_{y_0}}{n!} \qquad (23.91)$$

Formula (24.6) se numeşte *formula de interpolare* a lui *Newton*[2].

[2]Isac Newton (1642-1727) - "*Principiile matematice ale filosofiei naturale*", Ed. Academiei, Bucureşti, 1956

Scanned with OKEN Scanner

24. STUDIUL VARIAŢIILOR SEZONIERE, CICLICE ŞI A CELOR ALEATOARE

24.1. Variaţiile sezoniere

Variaţiile sezoniere sunt fluctuaţii periodice subanuale, mai mult sau mai puţin regulate, care se suprapun peste variaţiile extrasezoniere. Perioada lor de manifestare poate fi *ziua* (consumul casnic orar de energie electrică, traficul orar de călători în oraşe, intensitatea convorbirilor telefonice pe orele unei zile etc.), *săptămâna* (durata zilelor efectiv lucrate pe săptămână), *luna* (producţia industrială pe luni, volumul vânzărilor de mărfuri dintr-un magazin pe luni), *trimestrul* (producţia trimestrială din agricultură, construcţii) etc.

Variaţiile sezoniere sunt determinate de factori multipli: climaterici, mod de viaţă, obiceiuri, tradiţii, dispoziţii legale ale căror efecte se produc cu regularitate la o dată fixă.

Printre cele mai importante cauze s-ar putea cita:

1) *Factorii climaterici*, care constituie cauza principală a variaţiilor sezoniere (vara este sezonul preferat pentru concedii, temperatura influenţează consumul casnic de energie, activitatea în construcţii, dar mai ales în agricultură).

2) *Durata inegală a lunilor* în zile de lucru, existenţa sărbătorilor legale.

3) *Periodicitatea cererii şi a ofertei* anumitor produse, mai ales a celor agricole, vânzările de la sfârşitul fiecărui an, cererea de automobile primăvara etc.

4) *Concediile de odihnă* ale salariaţilor, care în fiecare an sunt luate mai ales vara duc la o diminuare a activităţii de producţie şi a altor indicatori economici, în acest anotimp.

În analiza sezonalităţii fenomenelor se pot distinge mai multe situaţii în funcţie de caracterul staţionar sau evolutiv al tendinţei centrale şi de amplitudinea constantă sau variabilă a variaţiilor sezoniere. (fig. 24.1.)

În vederea aplicării metodelor de măsurare a sezonalităţii fenomenelor în timp, informaţiile iniţiale se pot sistematiza sub forma unui tabel cu dublă intrare, în care pe linii se înscriu termenii seriei dinamice corespunzători perioadelor subanuale din toţi anii, iar pe coloane termenii aceleiaşi serii corespunzători perioadelor subanuale din fiecare an (tabelul 24.1.).

Scanned with OKEN Scanner

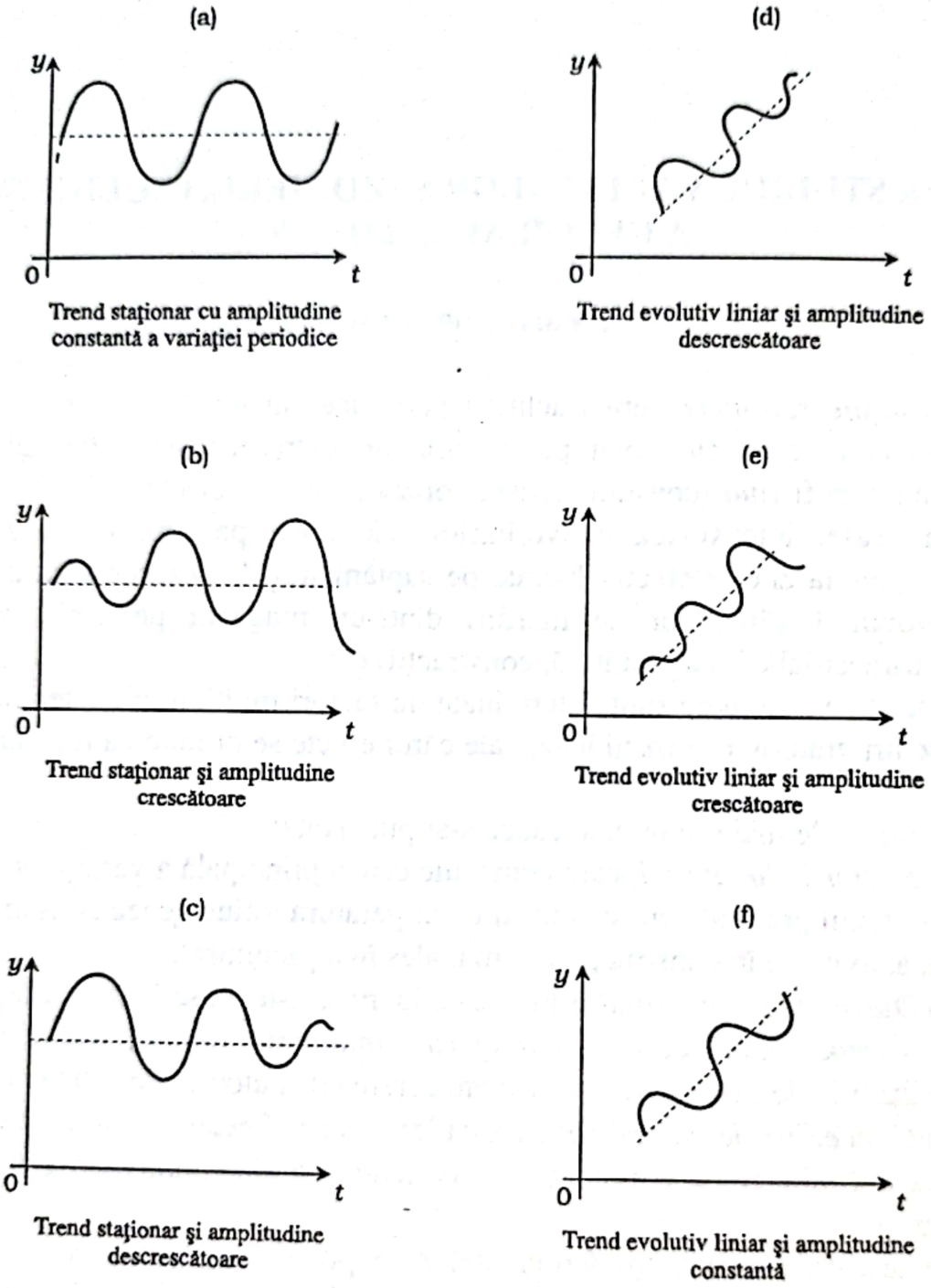
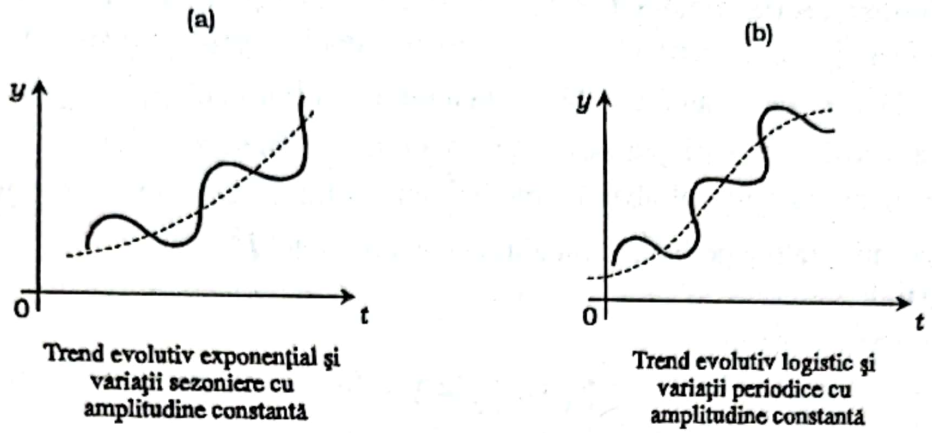


Figura. 24.1. Tipuri şi forme ale variaţiilor sezoniere în cazul unui trend liniar

Scanned with OKEN Scanner



Trend evolutiv exponenţial şi variaţii sezoniere cu amplitudine constantă

Trend evolutiv logistic şi variaţii periodice cu amplitudine constantă

Figura. 24.2. Tipuri şi forme ale variaţiilor sezoniere în cazul unui trend exponenţial

Tabelul 24.1. Prezentarea unei serii dinamice în vederea studierii sezonalităţii

Anul / Luna	1	2	...	j	...	p	Total
1	y_{11}	y_{12}	...	y_{1j}	...	y_{1p}	$y_{1.}$
2	y_{21}	y_{22}	...	y_{2j}	...	y_{2p}	$y_{2.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
i	y_{i1}	y_{i2}	...	y_{ij}	...	y_{ip}	$y_{i.}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	y_{k1}	y_{k2}	...	y_{kj}	...	y_{kp}	$y_{k.}$
Total	$y_{.1}$	$y_{.2}$	...	$y_{.j}$	...	$y_{.p}$	$y_{..}$
Medii lunare pe an	$\hat{y}_1$	$\hat{y}_2$	...	$\hat{y}_j$	...	$\hat{y}_p$	

În acest scop notăm termenii seriei cu y_{ij} în care i reprezintă perioada subanuală:

$i = \overline{1,2}$ pentru semestre; $i = \overline{1,12}$ pentru luni;

Scanned with OKEN Scanner

$i = \overline{1,4}$ pentru trimestre; $i = \overline{1,52}$ pentru săptămâni etc.

Termenii seriei dinamice se pot referi la o dată t precisă a fiecărei luni: $t = p(i-1) + j$ în particular în cazul unei serii lunare $(p = 2)$, iar $t = 12(i-1) + j$, deci valoarea observată a lui y va fi notată y_{ij} sau y_t.

Caracterizarea variaţiei sezoniere a unui fenomen se poate efectua atât în mărimi absolute, cu ajutorul abaterii mediei sezonului $\bar{y}_i$ faţă de media generală $\bar{y}$, cât şi în mărimi relative pe baza indicelui de sezonalitate I^S.

Astfel:

- mărimi absolute:

$$\Delta_{y_i} = \bar{y}_i - \bar{y}$$

$$\Delta_{y_i} > 0 \quad \text{sau} \quad \Delta_{y_i} = 0 \quad \text{sau} \quad \Delta_{y_i} < 0$$

- în mărimi relative:

$$I^s = \frac{\bar{y}_i}{\bar{y}} \cdot 100$$

$$I^S > 0 \quad \text{sau} \quad I^S = 100 \quad \text{sau} \quad I^S < 100$$

1) *Metoda mediilor aritmetice* se poate utiliza în practică pentru caracterizarea sezonalităţii în cazul seriilor dinamice care prezintă un trend staţionar.

Într-o serie dinamică, pe luni calendaristice se pot calcula:

- totalurile pe linii:

$$y_{1.} = \sum_{j=1}^{p} y_{1j} \qquad y_{2.} = \sum_{j=1}^{p} y_{2j} \quad \cdots \quad y_{i.} = \sum_{j=1}^{p} y_{ij} \quad \cdots \quad y_{k.} = \sum_{j=1}^{p} y_{kj}$$

- totalurile pe coloane:

$$y_{.1} = \sum_{i=1}^{k} y_{i1} \qquad y_{.2} = \sum_{i=1}^{k} y_{i1} \quad \cdots \quad y_{.j} = \sum_{i=1}^{k} y_{ij} \quad \cdots \quad y_{.p} = \sum_{i=1}^{k} y_{ip}$$

- total general:

Scanned with OKEN Scanner

$$y_{..} = \sum_{i=1}^{k} y_{i.} = \sum_{j=1}^{p} y_{.j} = \sum_{i=1}^{k}\sum_{j=1}^{p} y_{ij}$$

• *Media generală* $\bar{y}$ lunară în perioada celor p ani:

$$\bar{y} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{p} y_{ij}}{k \cdot p}$$

• *Mediile lunare* $\bar{y}_i$, calculate pe fiecare lună i din cei p ani:

$$\bar{y}_1 = \frac{y_{1.}}{p}; \bar{y}_2 = \frac{y_{2.}}{p}; \cdots; \bar{y}_i = \frac{y_{i.}}{p}; \cdots; \bar{y}_k = \frac{y_{k.}}{p}$$

• *Indicii de sezonalitate* se obţin prin compararea sub formă de raport între medii lunare şi media generală:

$$I_i^s = \frac{\bar{y}_i}{\bar{y}} \cdot 100 \qquad i = \overline{1,k}$$

2) *Metoda trendului*

Indicele sezonalităţii se poate determina prin raportarea mediilor lunare pe fiecare an din cadrul perioadei la nivelurile lor obţinute printr-o ajustare analitică, după algoritmul de calcul:

• se determină mediile lunare pe fiecare an $\hat{y}_j$:

$$\hat{y}_j = \frac{y_{.j}}{k} \qquad j = \overline{1,p}$$

• se ajustează seria mediilor lunare $\hat{y}_j$, cu ajutorul metodei celor mai mici pătrate, obţinându-se nivelurile lunare $\tilde{y}_j$ ajustate;

• se determină indicii sezonalităţii pe baza formulei:

$$I^s = \frac{\hat{y}_j}{\tilde{y}_j} \cdot 100$$

Scanned with OKEN Scanner

24.2. Măsurarea variaţiilor sezoniere

24.2.1. Analiza sezonalităţii în cazul modelului multiplicativ

Modelul multiplicativ, definit de relaţia:

$$Y = T \cdot S \cdot C \cdot I$$

se aplică pentru măsurarea fluctuaţiilor sezoniere metoda *ratios-to-moving averages* astfel:

1) Înlăturarea (separarea) variaţiilor sezoniere şi a celor aleatoare:

$$T \cdot C = \frac{T \cdot S \cdot C \cdot I}{S \cdot I}$$

2) Separarea trendului şi a variaţiilor ciclice:

$$S \cdot I = \frac{T \cdot S \cdot C \cdot I}{T \cdot C}$$

3) Separarea fluctuaţiilor aleatoare I din S.I:

$$S = \frac{S \cdot I}{I}$$

În cazul în care Y reprezintă indicatori trimestriali, mediile mobile trimestriale centrate vor cuprinde numai tendinţa T şi variaţiile ciclice C, adică $T \cdot C$. Eliminând aceste două componente din modelul multiplicativ rezultă componenta sezonieră însoţită de fluctuaţiile aleatoare (cea mai mare parte a acestora):

$$Y = T \cdot C \cdot S \cdot I$$

$$\frac{Y}{T \cdot C} = \frac{T \cdot S \cdot C \cdot I}{T \cdot C} = S \cdot I$$

Scanned with OKEN Scanner

Tabelul 24.2. Calculul indicilor specifici de sezonalitate (Date convenţionale)

Anul	Trimestrul	Mii lei Y	Medii mobile trimestriale	Medii mobile centrate ($T.C$)	(%) faţă de mediile mobile ($Y/T.C$)
1981	I	630			
	II	604	677,25		
	III	690	685,75	681,50	101,25
	IV	785	699,25	692,50	113,36
1982	I	664	705,75	702,50	94,52
	II	658	709,25	707,50	93,00
	III	716	719,00	714,13	100,25
	IV	799	725,50	722,25	110,63
1983	I	703	728,75	727,13	96,68
	II	684	761,75	745,25	91,78
	III	729	810,50	786,13	92,73
	IV	931	795,75	803,13	115,92
1984	I	898	778,00	786,88	114,12
	II	625	748,25	763,13	81,90
	III	658	691,75	720,00	91,39
	IV	812	707,75	699,75	116,04
1985	I	672	735,25	721,50	93,14
	II	689	727,50	731,38	94,21
	III	768	754,00	740,75	103,68
	IV	781	770,50	762,25	102,45
1986	I	778	792,50	781,50	99,55
	II	755	823,25	807,88	93,45
	III	856	831,50	827,38	103,46
	IV	904	841,25	836,38	108,08
1987	I	811	845,25	843,25	96,18
	II	794	856,00	850,63	93,34
	III	872	887,50	871,75	100,03
	IV	947	893,25	890,38	106,36
1988	I	937	899,75	896,50	104,52
	II	817	907,00	903,38	90,44
	III	898	903,75	905,38	99,18
	IV	976	920,75	912,25	106,99
1989	I	924	933,00	926,88	99,69
	II	885	947,50	940,25	94,12
	III	947	959,50	953,50	99,32
	IV	1034	970,75	965,13	107,14
1990	I	972	983,25	977,00	99,49
	II	930	997,50	990,38	93,90
	III	997			
	IV	1091			

Scanned with OKEN Scanner

Tabelul 24.3. Separarea fluctuaţiilor aleatoare din indicii specifici de sezonalitate

Trimestrele / Anii	I	II	III	IV	Total
1	2	3	4	5	6
1981	-	-	101,25	113,36	-
1982	94,52	93,00	100,26	110,63	-
1983	96,68	91,78	92,73	115,92	-
1984	114,12	81,90	91,39	116,04	-
1985	93,14	94,21	103,68	102,45	-
1986	99,55	93,45	103,46	108,08	-
1987	96,18	93,34	100,03	106,36	-
1988	104,52	90,44	99,18	106,99	-
1989	99,69	94,12	99,32	107,14	-
1990	99,49	93,90	-	-	-
Total	897,89	826,74	891,30	986,97	-
Media ($\bar{y}1$)	99,77	91,86	99,03	109,66	400,32
Medii ajustate	99,69	91,79	98,95	109,57	400,00
Mediana	99,49	93,34	100,03	108,08	400,94
Mediana ajustată	99,26	93,12	99,80	107,82	400,00
Total modificat	690,63	650,72	696,23	768,48	-
Medie modificată	98,66	92,96	99,46	109,78	400,86
Medie ajustată	98,45	92,76	99,25	109,54	400,00

Separarea componentei sezoniere S de fluctuaţiile aleatoare I se prezintă în tabelul 24.3., în care indicii specifici de sezonalitate din *col.6* a tabelului anterior se înscriu într-un tabel cu dublă intrare în care pe linii se trec anii, iar pe coloane, trimestrele sau lunile etc. Astfel, pe col.2 sunt înscrişi indicii de sezonalitate ai trimestrelor I în succesiunea anilor, pe col. 3 cei ai trimestrelor II ş.a.m.d.

Ipotetic, datele din aceste coloane, aferente fiecărui trimestru, reflectă numai variaţiile factorilor aleatori. Aceasta înseamnă că dacă factorii aleatori n-ar fi exercitat nici o influenţă asupra volumului trimestrial al vânzărilor, atunci indicii de sezonalitate pe fiecare trimestru ar trebui să aibă aceeaşi valoare în toţi anii luaţi în studiu. În plus, se presupune că influenţa factorilor aleatori într-un sens este egal probabilă cu influenţa lor în sens contrar, iar valoarea lor medie este egală cu zero.

Scanned with OKEN Scanner

În vederea separării componentei S de fluctuaţiile aleatoare I se determină mediile fiecărui trimestru pentru toţi anii în trei moduri diferite (vezi cele trei secţiuni din partea de jos a tabelului 24.3.).

Prima secţiune de pe rândul din mijloc cuprinde mediile aritmetice obişnuite $\bar{y}$ a căror sumă este 400,32, deci puţin mai mare de 400. Aceste valori se ajustează pentru a se obţine suma lor egală cu 400, prin înmulţirea fiecărui indice cu raportul 400/400,32. Valorile mediei ajustate ale indicilor de sezonalitate, de pe ultimul rând al primei secţiuni, vor arăta proporţia vânzărilor din trimestrul respectiv faţă de media anuală echivalată cu 100%. Astfel, în trimestrul I al anilor studiaţi, volumul vânzărilor a fost mai mic cu 0,31%, în trimestrul II cu 8,26% şi în trimestrul III cu 1,05% faţă de media trimestrială anuală, în timp ce în trimestrul IV vânzările au depăşit media cu 9,57%.

În secţiunea a doua indicatorul tendinţei centrale este estimat cu ajutorul medianei, mai puţin sensibilă la valorile extreme.

Se observă că indicii sezonalităţii sunt diferiţi în cele două secţiuni ale tabelului. Diferenţa se datorează metodelor de calcul diferite. În mod obişnuit sunt preferaţi indicii din secţiunea doi, calculaţi cu ajutorul medianei, deoarece ei elimină din calcul fluctuaţiile aleatoare extreme, în timp ce indicii din secţiunea I iau în calcul aceste valori.

Pentru înlăturarea fluctuaţiilor aleatoare extreme se procedează, în secţiunea III a tabelului, la înlăturarea celor două valori extreme din fiecare trimestru şi apoi se determină mediile aritmetice ale valorilor indicilor de sezonalitate rămaşi, obţinând astfel valorile medii modificate ale indicilor sezonalităţii pe rândul din mijloc, valori care apoi se ajustează pe ultimul rând pentru ca suma lor să fie egală cu 400. Modelul de sezonalitate pentru o serie dinamică se poate menţine neschimbat sau se poate modifica în timpul perioadei de analiză ca urmare a influenţei exercitate de factorii sezonieri. Modificarea modelului de sezonalitate poate apare brusc sau în mod gradual şi aceasta determină modificarea indicilor de sezonalitate.

În cazuri concrete, pentru a verifica dacă modelul de sezonalitate se menţine neschimbat sau se modifică în perioada reflectată de o serie dinamică, se poate proceda la reprezentarea grafică a indicilor de sezonalitate specifici pentru fiecare trimestru sau lună etc. Dacă indicii de sezonalitate tind să fluctueze în jurul tendinţei lor centrale determinate prin metoda celor mai mici pătrate (trimestriale), fără a se putea desprinde vreo tendinţă de creştere sau de scădere, atunci modelul de sezonalitate se consideră invariabil, iar în cazul contrar modelul de sezonalitate este variabil.

Scanned with OKEN Scanner

Tabelul 24.4. Ajustarea indicilor de sezonalitate aferenţi trimestrului I

Anul	%	t_i	t_i^2	$t_i y_i$	$y_{t_i} = a + bt_i$
1981	-	-	-	-	98,695
1982	94,5	-4	16	-378,0	98,910
1983	96,7	-3	9	-290,1	99,125
1984	114,1	-2	4	-228,2	99,340
1985	93,1	-1	1	-93,1	99,555
1986	99,6	0	0	0	99,770
1987	96,2	1	1	96,2	99,985
1988	104,5	2	4	209,0	100,200
1989	99,7	3	9	299,1	100,415
19902	99,53	4	16	398,0	100,630
Total	897,9	0	60	12,9	897,930

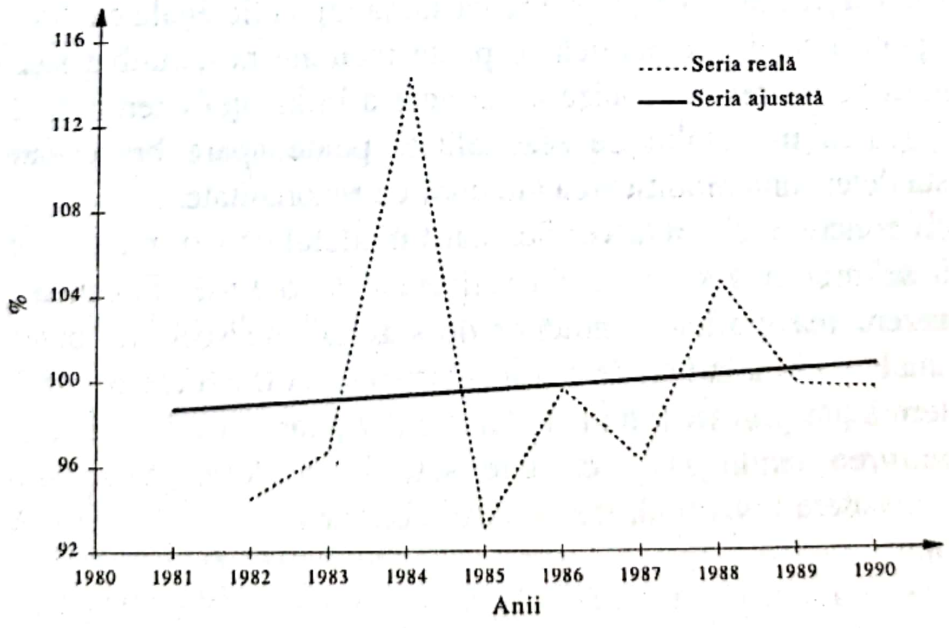


Figura 24.3. Indicii de sezonalitate pentru trimestrul I

Scanned with OKEN Scanner

Parametrii ecuaţiei liniare de trend pentru trimestrul I vor fi:

$$a_I = \frac{897,9}{9} = 99,77$$

$$b_I = \frac{12,9}{60} = 0,215$$

Tabelul 24.5. Ajustarea indicilor de sezonalitate aferenţi trimestrului II

Anul	%	t_i	t_i^2	$t_i y_i$	$y_{t_i} = a + bt_i$
1981	-	-	-	-	89,58
1982	93,0	-4	16	-327,0	90,02
1983	91,8	-3	9	-275,4	90,16
1984	81,9	-2	4	-163,8	90,90
1985	94,2	-1	1	-94,2	91,34
1986	93,5	0	0	0	91,78
1987	93,3	1	1	93,3	92,22
1988	90,4	2	4	180,8	92,66
1989	94,1	3	9	282,9	93,10
1990	93,9	44	16	375,65	93,54
Total	826,1	0	60	26,6	826,2

Parametrii ecuaţiei liniare de trend pentru trimestrul II vor fi:

$$\begin{cases} a_{II} = \frac{826,1}{9} = 91,78 \\ b_{II} = \frac{26,6}{60} = 0,44 \end{cases}$$

Scanned with OKEN Scanner

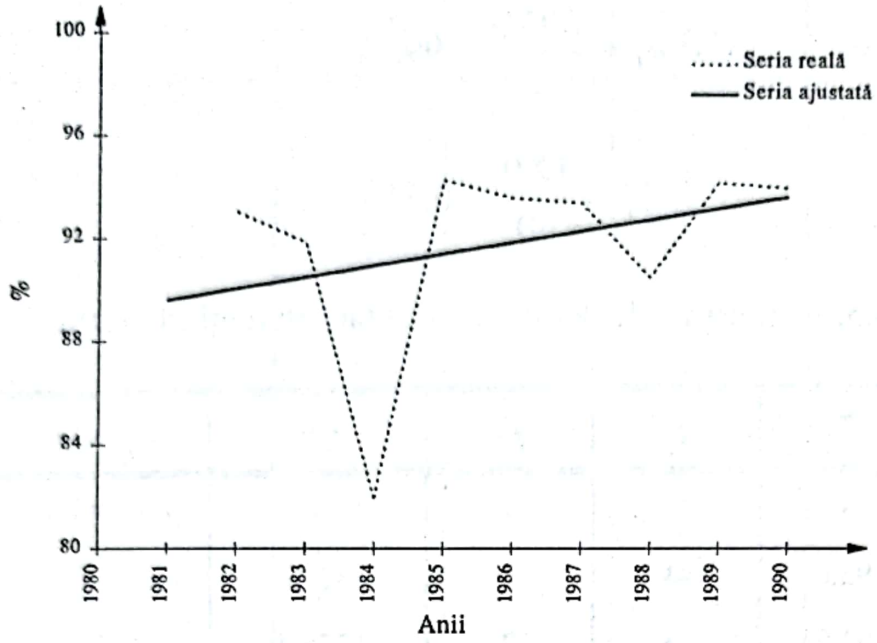


Figura 24.4. Indicii de sezonalitate pentru trimestrul al II-lea

Tabelul 24.6. Ajustarea indicilor de sezonalitate aferenţi trimestrului III

Anul	%	t_i	t_i^2	$t_i y_i$	$y_{t_i} = a + bt_i$
1981	101,3	-4	16	-405,2	98,02
1982	100,3	-3	9	-300,9	98,27
1983	92,7	-2	4	-185,4	98,53
1984	91,4	-1	1	-91,4	98,78
1985	103,7	0	0	0	99,30
1986	103,5	1	1	103,5	99,55
1987	100,0	2	4	200,0	99,81
1988	99,2	3	9	297,6	100,06
1989	99,3	4	16	397,2	100,32
1990	-	-	-	-	
Total	891,4	0	60	15,4	891,36

Scanned with OKEN Scanner

Parametrii ecuaţiei liniare corespunzători trimestrului III se prezintă astfel:

$$\begin{cases} a_{III} = \dfrac{891,4}{9} = 99,04 \\ b_{III} = \dfrac{15,4}{60} 0,256 \end{cases}$$

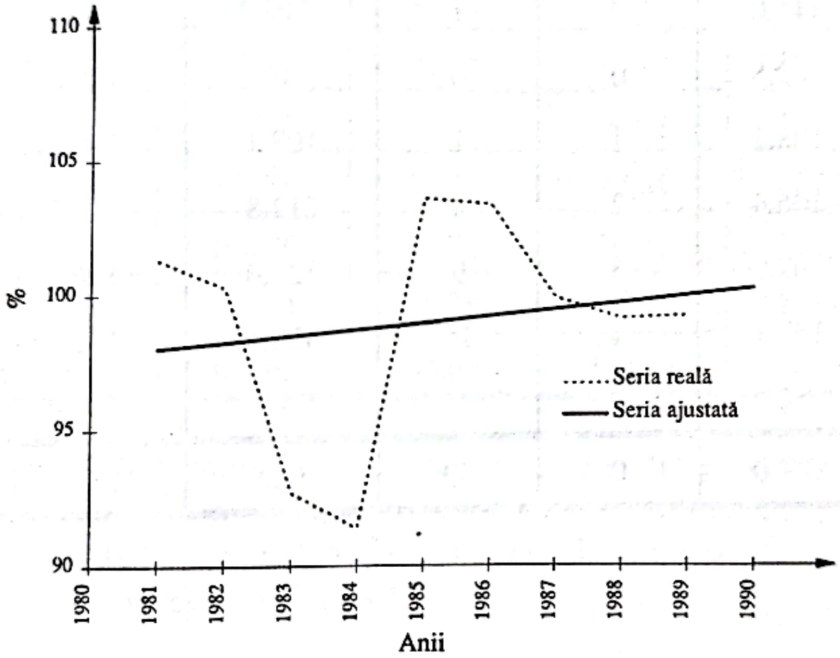


Figura 24.5 Indicii de sezonalitate pentru trimestrul al III-lea

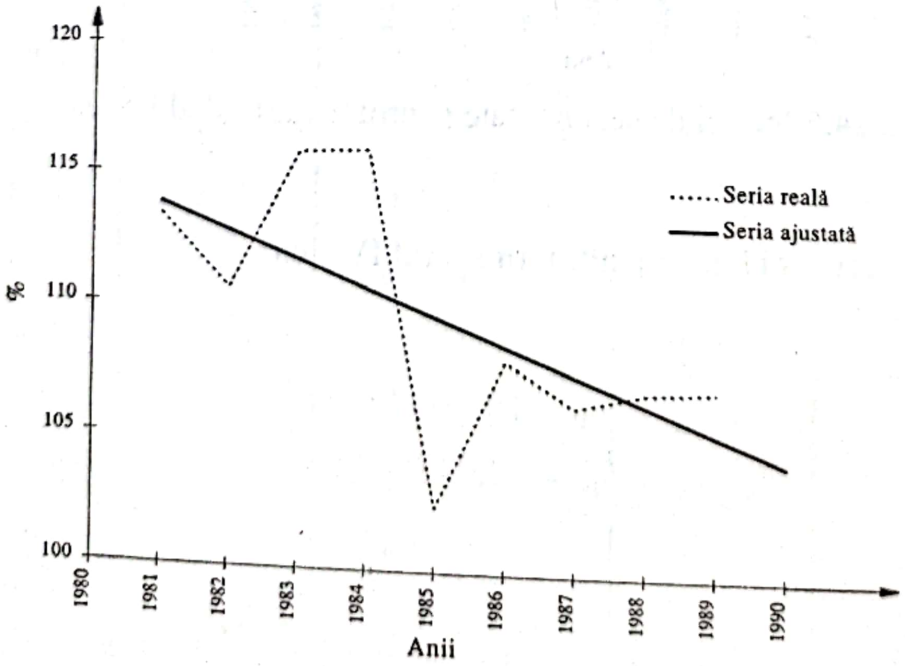
Parametrii ecuaţiei liniare pentru trimestrul IV sunt:

$$\begin{cases} a_{IV} = 109,667 \\ b_{IV} = -1,048 \end{cases}$$

Scanned with OKEN Scanner

Tabelul 24.7. Ajustarea indicilor de sezonalitate aferenţi trimestrului IV

Anul	%	t_i	t_i^2	$t_i y_i$	$y_{t_i} = a + bt_i$
1981	113,4	-4	16	-453,6	113,859
1982	110,6	-3	9	-331,8	112,811
1983	115,9	-2	4	-231,8	111,763
1984	116,0	-1	1	-116,0	110,715
1985	102,5	0	0	0	109,667
1986	108,1	1	1	108,1	108,619
1987	106,4	2	4	212,8	107,571
1988	107,0	3	9	321,0	106,523
1989	107,1	4	16	428,4	105,475
1990	-6	-	-	-	104,427
Total	987,0	0	60	-62,9	987,003



Scanned with OKEN Scanner

Figura 24.6. Indicele de sezonalitate pentru trimestrul al IV-lea

Tabelul 24.8. Indicii de sezonalitate ajustaţi

Trimestrul / Anul	*I*	*II*	*III*	*IV*	Total
1981	97,80	89,55	98,02	113,86	400,13
1982	98,91	90,00	98,29	112,41	399,61
1983	99,13	90,45	98,56	111,76	399,90
1984	99,34	90,90	98,83	110,72	399,79
1985	99,56	91,35	99,10	109,67	399,68
1986	99,77	91,80	99,37	108,62	399,56
1987	99,99	92,25	99,64	107,57	399,45
1988	100,20	92,70	99,91	106,52	399,33
1989	100,41	93,15	100,18	105,48	399,22
1990	100,63	93,60	100,45	104,43	399,11

Din tabelele 24.4., 24.5., 24.6., 24.7., ca şi din graficele 24.3., 24.4., 24.5., 24.6., se constată că modelul sezonier liniar în primele trei trimestre este uşor crescător (b_I = 0,215; b_{II} = 0,45; b_{III} = 0,27), în timp ce în cel de-al patrulea se modifică liniar descrescător (b_{IV} = - 1,048).

24.2.2. Analiza sezonalităţii în cazul modelului aditiv

În condiţiile modelului aditiv de agregare a componentelor variaţiei în timp a fenomenelor se porneşte de la aproximarea tendinţei centrale T cu ajutorul mediilor mobile.

Pentru ilustrarea metodei de analiză considerăm seria dinamică din tabelul 24.9 referitoare la evoluţia volumului vânzărilor unui magazin pe trimestre în decursul unui deceniu. Mediile mobile estimează nu numai tendinţa de lungă durată ci şi componenta ciclică C. Deoarece seria luată în calcul se referă la o perioadă scurtă de timp (doar un deceniu), nu se poate argumenta ipoteza existenţei unei variaţii ciclice. De aceea modelul aditiv obişnuit $Y = T + C + S + I$ se va simplifica prin eliminarea lui C şi va deveni $Y = T + S + I$.

Scanned with OKEN Scanner

Tabelul 24.9. Analiza sezonalităţii în cazul modelului aditiv simplificat: $Y = T + S + I$

Anul	Trim.	Volumul vânzării $Y_i = T + S + I$	Medii Mobile centrate T	Serii Detrendizate $Y_i - T = S + I$	Componenta sezonieră	Seria sezonieră ajustată $Y - S = T + I$	Componenta reziduală $Y - S - T = I$
1981	I	630	-	-	- 0,97	630,97	-
	II	604	-	-	-67,19	671,19	-
	III	690	682	8	- 7,75	697,75	15,75
	IV	785	693	92	75,92	709,08	16,08
1982	I	664	703	- 39	- 0,97	664,97	- 38,03
	II	658	708	- 50	-67,19	725,19	17,19
	III	716	714	2	- 7,75	723,75	9,75
	IV	799	722	77	75,92	723,08	1,08
1983	I	703	727	- 24	- 0,97	703,97	- 23,03
	II	684	745	- 61	-67,19	751,19	6,19
	III	729	786	- 57	- 7,75	736,75	- 49,25
	IV	931	803	128	75,92	855,08	52,08
1984	I	898	787	111	- 0,97	898,97	111,97
	II	625	763	-138	-67,19	692,19	- 70,81
	III	658	720	- 62	- 7,75	665,75	- 54,25
	IV	812	700	112	75,92	736,08	36,08
1985	I	672	722	- 50	- 0,97	672,97	- 49,03
	II	689	731	- 42	-67,19	756,19	25,19
	III	768	741	27	- 7,75	775,75	34,75
	IV	781	762	19	75,92	705,08	- 56,92
1986	I	778	782	- 4	- 0,97	778,97	- 3,03
	II	755	808	- 53	-67,19	822,19	14,19
	III	856	827	29	- 7,75	863,75	36,75
	IV	904	836	68	75,92	828,08	- 7,92
1987	I	811	843	- 32	- 0,97	811,97	- 31,03
	II	794	851	- 57	-67,19	861,19	10,19
	III	872	872	0	- 7,75	879,75	7,75
	IV	947	890	57	75,92	871,08	- 18,92
1988	I	937	897	40	- 0,97	937,97	40,97
	II	817	903	- 86	-67,19	884,19	- 18,81
	III	898	905	- 7	- 7,75	905,75	0,75
	IV	976	912	64	75,92	900,08	- 11,92
1989	I	924	927	- 3	- 0,97	924,97	- 2,03
	II	885	940	- 55	-67,19	952,19	12,19
	III	947	954	- 7	- 7,75	954,75	0,75
	IV	1034	965	69	75,92	958,08	- 6,92
1990	I	972	977	- 5	- 0,97	972,97	- 4,03
	II	930	990	- 60	-67,19	997,19	7,19
	III	997	-	-	- 7,75	1004,75	-
	IV	1091	-	-	75,92	1015,08	-

Sursa: *Date convenţionale*

Scanned with OKEN Scanner

În aceste condiţii, scăzând din termenii seriei iniţiale Y componenta tendenţială T se va obţine componenta sezonieră S şi cea aleatoare I (col. 5 a tabelului 24.9) : $Y - T = S + I$.

Membrul al doilea al acestei relaţii poartă denumirea de *serie detrendizată.*

În vederea separării componentei sezoniere din termenii seriei detrendizate se determină valorile medii trimestriale (tabelul 24.10).

Se observă că suma valorilor termenilor seriei detrendizate este egală cu 11, deci diferită de zero.

Dacă mediile mobile centrate ar fi descris în mod complet şi adecvat tendinţa centrală de evoluţie a volumului vânzărilor, atunci ar fi trebuit ca scăderile vânzărilor din primele trei trimestre să fie total compensate de creşterea acestora în trimestrul patru: $(-6)+(-602)+(-67)+686=11\neq 0$. Corectarea datelor de pe primul rând de sub tabel se va face prin scăderea din fiecare tabel trimestrial a mărimii $\frac{11}{4}=2{,}75$.

Împărţind aceste valori corectate la numărul anilor (9) corespunzători fiecărui trimestru din serie se obţine valoarea medie a *componentei sezoniere*, (vezi ultimul rând al tabelului 24.10). Rezultă că, drept urmare a influenţei factorilor sezonieri, volumul vânzărilor s-a redus în primele trei trimestre ale anilor; în cea mai mare măsură în trimestrul doi, reduceri compensate de creşterea vânzărilor înregistrată în trimestrul al patrulea.

Tabelul 24.10. Seria detrendizată

Trimestrul / Anul	*I*	*II*	*III*	*IV*	Total
1981	-	-	8	92	
1982	-39	-50	2	77	
1983	-24	-61	-57	128	
1984	111	-138	-62	112	
1985	-50	-42	27	19	
1986	-4	-53	29	68	
1987	-32	-57	0	57	
1988	40	-86	-7	64	
1989	-3	-55	-7	69	
1990	-5	-60	-	-	
Total	-6	-602	-67	686	11
Total corectat (-2,75)	-8,75	-604,75	-69,75	683,25	0
(Media trim.):9	-0,97	-67,19	-7,75	75,92	

Scanned with OKEN Scanner

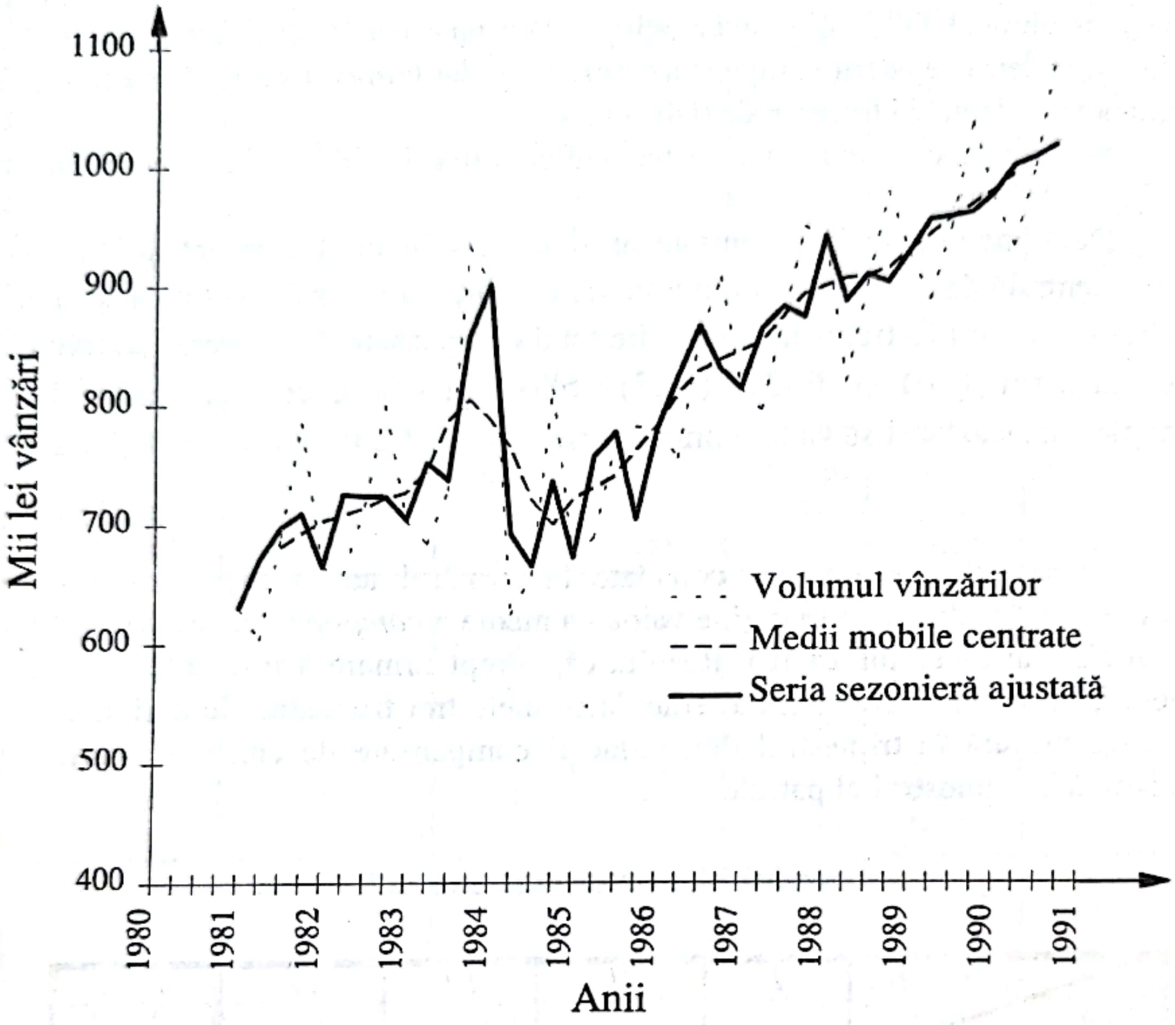


Figura 24.7. Seria sezonieră ajustată

Scăzând din seria detrendizată componenta sezonieră S rezultă mărimea componentei aleatoare I (col. 6 a tabelului 24.9).

Analiza seriei dinamice, pe baza modelului aditiv, se poate completa în mod sugestiv cu ajutorul reprezentărilor grafice atât a seriei dinamice iniţiale, cât şi a componentelor sale (fig.24.7).

Scanned with OKEN Scanner

24.3. Variaţiile ciclice

În jurul tendinţei pe termen lung pot avea loc fluctuaţii determinate de acţiunea unor factori conjuncturali, cum ar fi fazele ciclului economic: expansiunea, criza, depresiunea şi avântul. Fiecare fază a ciclului economic caracterizează o anumită stare a economiei şi îndeplineşte o anumită funcţie specifică.

Variaţiile conjuncturale sau extrasezoniere cuprind *variaţiile ciclice* şi *trendul.*

Trendul şi variaţiile sezoniere reprezintă singurele componente ale evoluţiei fenomenelor în timp care pot fi evidenţiate şi măsurate cu un anumit grad de precizie şi care prezintă stabilitate acceptabilă şi constanţă ca model în timp.

Variaţiile ciclice se întind repetabil pe perioade mai mari de un an, ale căror sincronizare şi amplitudine sunt departe de a fi regulate şi predictibile. Ele se apropie de forma unor curbe oscilatorii, cum ar fi sinusoida.

Informaţiile deduse din evoluţia anterioară a fenomenelor cu variaţie ciclică nu sunt suficiente pentru a explica această componentă a seriilor dinamice şi de aceea în practică se recurge la utilizarea unor modele statistico-economice mai complexe numite *modele econometrice.*

Chiar dacă nu pot fi prognozate, fluctuaţiile ciclice pot fi separate de alte componente ale variaţiei cu ajutorul *indicilor ciclicităţii* care servesc pentru modificarea tendinţei şi a componentei sezoniere în calculele de prognoză.

În separarea variaţiilor ciclice, se disting două situaţii în funcţie de situaţia în care indicatorii seriilor dinamice se referă la ani sau la perioade subanuale.

1. În cazul *datelor anuale*, care cuprind numai două componente (trend şi componenta ciclică) însoţite probabil şi de o parte din componenta aleatoare, în condiţiile:

Modelului multiplicativ:

$$Y = T \cdot C$$

sau, mai precis,

$$Y = T \cdot C \cdot I$$

Modelului aditiv:

$$Y = T + C$$

$$Y = T + C + I$$

de unde componenta ciclică va fi:

Scanned with OKEN Scanner

$$\frac{Y}{Y_t}=\frac{TC}{T}=C \qquad Y-T=C$$

sau

$$\frac{Y}{Y_t}=\frac{T\cdot C\cdot I}{T}=C\cdot I \qquad Y-T=C+I$$

Seriile dinamice din care a fost separată (eliminată) componenta trend se numesc *serii dinamice (de timp) staţionare*.

2. În cazul *datelor pe perioade subanuale* (pe trimestre), separarea componentei ciclice pentru modelul multiplicativ va fi:

$$\frac{Y}{S\cdot Y_t}=\frac{T\cdot C\cdot S\cdot I}{T\cdot S}=C\cdot I$$

Ţinând seama de faptul că indicii de sezonalitate reflectă şi cea mai mare parte a componentei aleatoare ($S\cdot I$) rezultă:

$$\frac{Y}{Y_t(S\cdot I)}=\frac{T\cdot C\cdot S\cdot I}{T\cdot(S\cdot I)}=C$$

În mod analog, pentru modelul aditiv: $Y-T-S=C+I$

de unde $Y-T-S-I=C$

Parametrii trendului liniar vor fi:

$$a=\frac{\sum_{i=1}^{n}Y_i}{n}=\frac{3261}{10}=3262{,}1$$

Scanned with OKEN Scanner

$$a = \frac{\sum_{i=1}^{n} t_i Y_i}{\sum_{i=1}^{n} t_i^2} = \frac{2271}{330} = 69{,}03$$

Tabelul 24.11. Volumul vânzărilor anuale

Anul	Y_i	t_i	t_i^2	$t_i y_i$	$Y_t = a + bt$	$\frac{Y_i}{T} \cdot 100$
1981	2709	-9	81	-24381	2917	92,87
1982	2837	-7	49	-19859	2986	95,01
1983	3047	-5	25	-15235	3055	99,74
1984	2993	-3	9	-8979	3124	95,81
1985	2910	-1	1	2910	3193	91,14
1986	3293	1	1	3293	3331	98,86
1987	3424	3	9	10272	3400	100,71
1988	3628	5	25	18140	3469	104,58
1989	3790	7	49	26530	3538	107,12
1990	3990	9	81	35910	3607	110,62
Total	32621	0	330	22781	32620	

Deci $y_t = 3262{,}1 + 69{,}03\, t_i$

Componenta ciclică se va obţine prin raportarea termenilor reali ai seriei y_i (col. 1 a tabelului 24.11) la valorile trendului y_t (col.5 din tabelul 24.11).

24.4. Variaţiile aleatoare

În jurul primelor trei componente ale variaţiei au loc şi *fluctuaţii aleatoare*, care au, în general, o mică amplitudine, determinate fie de cauze necunoscute, fie de cauze ocazionale: greve, dificultăţi financiare, modificări ale legislaţiei economice etc. Se mai numesc şi *fluctuaţii reziduale*, deoarece nu pot fi luate în considerare în primele trei componente ale variaţiei.

Scanned with OKEN Scanner

Tabelul 24.12. Ajustarea liniară a volumului vânzărilor

Anul	Trim.	Y	t_i	t_i^2	t_iY_i	Y_i
1981	*I*	630	-39	1521	-24570	641,7
	II	604	-37	1369	-22348	650,6
	III	690	-35	1225	-24150	659,5
	IV	785	-33	1089	-25905	668,4
1982	*I*	664	-31	961	-20584	677,4
	II	658	-29	841	-19082	686,3
	III	716	-27	729	-19332	695,2
	IV	799	-25	625	-19975	704,1
1983	*I*	703	-23	529	-16169	713,0
	II	684	-21	441	-14364	721,9
	III	729	-19	361	-13851	730,8
	IV	931	-17	289	-15827	739,8
1984	*I*	898	-15	225	-13470	748,7
	II	625	-13	169	- 8125	757,6
	III	658	-11	121	- 7238	766,5
	IV	812	- 9	81	- 7308	775,4
1985	*I*	672	- 7	49	- 4704	784,3
	II	689	- 5	25	- 3445	793,2
	III	768	- 3	9	- 2304	802,2
	IV	781	- 1	1	- 781	811,1
1986	*I*	778	1	1	778	820,0
	II	755	3	9	2265	828,9
	III	856	5	25	4280	837,8
	IV	904	7	49	6328	846,7
1987	*I*	811	9	81	7299	855,6
	II	794	11	121	8734	864,6
	III	872	13	169	11336	873,5
	IV	947	15	225	14205	882,4
1988	*I*	937	17	289	15929	891,3
	II	817	19	361	15523	900,2
	III	898	21	441	18858	909,1
	IV	976	23	529	22448	918,0
1989	*I*	924	25	625	23100	927,0
	II	885	27	729	23895	935,9
	III	947	29	841	27463	944,8
	IV	1034	31	961	32054	953,7
1990	*I*	972	33	1089	32076	962,6
	II	930	35	1225	32550	971,5
	III	997	37	1369	36889	980,4
	IV	1091	39	1521	42549	989,3
Total		32621		21320	95027	32621,0

Scanned with OKEN Scanner

VII. INDICII

Scanned with OKEN Scanner

25. INDICII

Scanned with OKEN Scanner

25. INDICI

25.1. Noţiunea şi importanţa indicilor

Studiul variaţiei fenomenelor, în general, şi a variaţiei lor în timp, în special, i-a preocupat pe oamenii de ştiinţă dintotdeauna, dar primele începuturi ale tratării lor ştiinţifice au apărut de abia spre mijlocul secolului al *XVIII-lea.*

Noţiunea de indice a apărut la începutul secolului al *XVIII^lea*, odată cu studiul întreprins de episcopul anglican *W. Fleetwood* (*Chronicon Preciosum*, Londra, 2^nd ed., 1745) asupra evoluţiei preţurilor în Anglia între anii 1440 şi 1700, având ca scop să arate dacă o clauză din statutul unui colegiu, stabilită cu trei secole în urmă, îşi mai păstra valabilitatea, ţinând cont de evoluţia preţurilor. Această clauză prevedea că toţi membrii unui colegiu trebuiau să jure că vor părăsi colegiul dacă averea lor va depăşi 5 £ pe an. În acest scop, el a luat în calcul patru produse de consum curent : grâu, carne, băutură şi îmbrăcăminte, cercetând evoluţia preţurilor acestora de-a lungul timpului. Concluzia sa a fost următoarea : cu 5 £ în perioada 1440 - 1460 se puteau cumpăra tot atâtea produse de consum (din cele alese de el), cât cu 30 £ în 1700.

Mai târziu, în anul 1738, *C. Dutot* a studiat scăderea puterii de cumpărare a monedei franceze în perioada cuprinsă între domniile a doi regi, *LOUIS al XII^lea* şi *LOUIS al XV^-lea* .Veniturile celor doi regi au fost:

- în anul 1515 *LOUIS al XII^lea* : 7 650 000 £;
- în anul 1735 *LOUIS al XV^lea* : 100 000 000 £.

Pentru a determina care dintre cei doi suverani avea un venit mai mare, el a luat în considerare mai multe preţuri unitare în cei doi ani pentru : un pui de găină, un iepure, un porumbel, valoarea unei zile de muncă etc., iar pentru comparaţia între cele două perioade, a folosit următoarea formulă[1]:

$$I^{p}_{t/0} = \frac{\sum_{i=1}^{n} p_{ti}}{\sum_{i=1}^{n} p_{0i}}$$

[1]C. Dutot: *Reflexions politiques sur les finances et le commerce*, The Hague, 1738.

Scanned with OKEN Scanner

Menţionăm că preţurile unitare însumate mai sus nu sunt ponderate. Concluzia sa a fost că veniturile lui *LOUIS al XII^lea* au fost mai mici decât cele ale lui *LOUIS al XV^lea*.

Fie p_{ti} preţurile unitare pentru produsele $i=\overline{1,n}$, în perioadele $t = 0, 1, 2, ..., k$, atunci indicele *Dutot* va avea următoarea expresie:

$$I^{p}_{t/0} = \frac{p_{t1} + p_{t2} + \cdots + p_{ti} + \cdots + p_{tn}}{p_{01} + p_{02} + \cdots + p_{0i} + \cdots + p_{0n}} = \frac{\sum_{i=1}^{n} p_{ti}}{\sum_{i=1}^{n} p_{0i}}$$

Un sfert de secol mai târziu, matematicianul şi astronomul italian *Gian Rinaldo Carli* [2] *(1720 - 1795)*, profesor la Universitatea din Padova, a construit un indice al preţurilor sub formă de medie aritmetică simplă a indicilor individuali, pentru a măsura scăderea puterii de cumpărare a banilor, ca urmare a importului de argint în *Europa* din *America*. *Carli* a considerat doar preţurile la trei produse principale (grâu, vin şi ulei) în anii *1500* şi *1750*, notate respectiv cu p_1, p_2 şi p_3 şi a calculat variaţia lor relativă în decursul celor *250* de ani consideraţi. Apoi a determinat media aritmetică simplă a celor trei mărimi relative, obţinând în felul acesta un indice al preţurilor care, după cum se ştie, se apropie de mărimea inversă a indicelui puterii de cumpărare a banilor:

$$I^{p}_{t/0} = \left(\frac{p_{11}}{p_{01}} + \frac{p_{12}}{p_{02}} + \frac{p_{13}}{p_{03}}\right) : 3$$

sau

$$I^{p}_{t/o} = \frac{1}{n}\sum_{i=1}^{n} \frac{p_{ti}}{p_{0i}}$$

Sfârşitul secolului al *XVIII^lea* este marcat de contribuţia lui *Sir George Shuckburgh Evelyn*, care în lucrarea " *An Account of Some Endeavours to Ascertain*

[2] Carli G.R., *Del valore e della proporzione de metalli monetati con i generi in Italia prima delle Scoperte dell Indice con confronto del valore e delle proporzione dei tempi nostri*, 1764, publicat de Custodi în: Scrittori italiani de economia politica, parte moderna, vol.XIII, p. 297-366.

Scanned with OKEN Scanner

a Standard of Weight and Measure "Phil. Trans. 113 (1798), a construit un indice al preţurilor pentru perioada 1550 - 1800, introducând noţiunea de an de bază (considerat de el anul1550). Pentru prima dată, alături de bunuri, el ia în studiu şi preţurile unor servicii.

La începutul secolului al *XX-lea* se semnalează contribuţia lui *Joseph Lowe*, considerat de M. Kendall drept " *părintele adevărat al indicilor* ", care în lucrarea *The Present State of England*, apărută la Londra în anul 1922, se ocupă de numeroase probleme ale construirii indicilor, introducând printre altele *noţiunea de pondere* a preţurilor din diferite perioade p_0 şi p_t, cu cantităţile din perioada iniţială q_0, obţinând astfel un *indice agregat* al preţurilor:

$$I_{t/0}^{p} = \frac{\sum_{i=0}^{n} p_{ti} q_{0i}}{\sum_{i=0}^{n} p_{0i} q_{0i}}, \qquad i = \overline{0,n}$$

Mai târziu, acest indice a devenit cunoscut drept indicele de tip *Laspeyres*.

În anul 1863 se impune *W. S. Jevons*, care utilizează conceptul de medie în constituirea indicilor, recomandând *media geometrică*[3].

În cea de-a doua jumătate a secolului al *XIX-lea* s-au impus contribuţiile a trei mari statisticieni germani : *E. Laspeyres, M. W. Drobisch* şi *H. Paasche*, care au studiat evoluţia preţurilor bunurilor materiale observate în oraşul Hamburg, susţinând împreună necesitatea utilizării *ponderilor* în construirea indicilor de grup şi respingând metodele de calcul care nu ţin seama de acest element.

R. H. I. Palgrave în anul 1886 propune formula indicelui mediu aritmetic ponderat al preţurilor individuale, folosind ca pondere valoarea totală a bunurilor, aferentă perioadei curente $v_1 = q_1 p_1$, propunând formula:

$$I_{t/0}^{p} = \frac{\sum_{i=0}^{n} i_{t/0}^{p} p_{ti} q_{ti}}{\sum_{i=0}^{n} p_{ti} q_{ti}}$$

[3] W.S.Jevons : *A Serious Fall in the Value of Gold Ascertained and its Social Effects Set Forth*, London, 1863.

Scanned with OKEN Scanner

Indicii reprezintă o categorie specială a indicatorilor statistici cu ajutorul cărora se măsoară variaţia medie relativă a fenomenelor sociale şi economice în timp şi în spaţiu sau în raport cu un alt nivel, considerat ca *nivel de bază* sau *etalon*.

Prin definiţie, *orice indice se obţine prin raportarea între ele a două niveluri absolute ale aceluiaşi fenomen sau grup de fenomene*. Unul dintre niveluri reprezintă volumul fenomenului din perioada care se raportează, iar celălalt, volumul fenomenului din perioada considerată ca bază de raportare. Nivelul ales ca bază de comparaţie trebuie să fie un *nivel normal* din dezvoltarea fenomenului cercetat, un nivel care să aibă o *semnificaţie deosebită* (să reprezinte o anumită etapă sau treaptă de dezvoltare), în aşa fel încât să justifice alegerea lui ca *etalon* în analiza statistică.

La construirea unui indice, trebuie să se ţină seama de următoarele reguli:

1) *Orice indice este rezultatul comparării a două situaţii sau a unei comparaţii binare*. Prin generalizare se poate ajunge la compararea în serie. Cele două situaţii din comparaţia binară pot fi două perioade de timp, două localităţi sau zone diferite, etc

2) *Informaţiile iniţiale* necesare comparaţiei binare, cu privire la toate bunurile economice luate în studiu, *trebuie să fie complete şi precise*. De exemplu, dacă se calculează un indice al preţurilor pentru o serie de produse este necesar să se cunoască cantităţile şi preţurile fiecărui produs în cele două momente sau perioade de comparat. Prin înmulţirea cantităţilor q_i cu preţurile p_i, pentru fiecare produs, se obţin valorile $v_i = q_i p_i$, iar prin împărţirea sumei valorilor tuturor produselor $\sum_{i=1}^{n} v_i = \sum_{i=1}^{n} q_i p_i$ din cele două perioade se obţine un indice a cărui valoare este, de regulă, diferită de unitate.

Deoarece indicii se obţin prin raportarea nivelului atins de unul sau mai multe fenomene în două momente, perioade sau zone diferite, ei *sunt indicatori fără dimensiune*, adică independenţi de unitatea de măsură a fenomenului luat în calcul.

25.2. Clasificarea indicilor

În teoria şi practica statistică se utilizează numeroşi indici care diferă între ei din multiple puncte de vedere. De aceea, pentru formarea unei imagini coerente, de

Scanned with OKEN Scanner

ansamblu, asupra lor este necesar să fie clasificaţi după criterii diferite, cum ar fi : sfera de cuprindere, natura caracteristicii, modul de alegere a bazei de comparaţie, procedeul de calcul, etc.

1. Din punct de vedere al *sferei de cuprindere*, indicii se împart în două categorii:

- *indici individuali, simpli* sau *elementari* care măsoară variaţia relativă a unor fenomene simple, singulare, în timp sau în spaţiu;
- *indici de grup, compuşi* sau *sintetici* care măsoară variaţia medie relativă a unor fenomene colective în raport cu o anumită bază de comparaţie.

2. După *natura caracteristicii* studiate, se disting două categorii de indici:

- indici ai variaţiei în timp sau *indici ai dinamicii*;
- indici ai variaţiei în spaţiu sau *indici teritoriali*;

3. În funcţie de *modul de alegere a bazei de raportare*, atât indicii individuali, cât şi cei de grup se împart în:

- *indici cu bază fixă*;
- *indici cu bază mobilă*.

4. Din punct de vedere al *ponderilor utilizate*, deosebim:

- *indici cu ponderi constante*;
- *indici cu ponderi variabile*.

5. Din punct de vedere al *procedeului de calcul*, indicii de grup se împart în:

- *indici agregaţi*;
- *indicii calculaţi ca medii aritmetice* ponderate ale indicilor individuali;
- *indicii calculaţi sub formă de medie armonică* ponderată a indicilor individuali;
- *indici calculaţi sub formă de medie geometrică*;
- *indici calculaţi prin raportarea a două medii aritmetice ponderate* etc.

25.3. Indicii dinamicii

Indicii dinamicii reprezintă categoria cea mai importantă şi mai dezvoltată din cadrul tuturor indicilor.

După cum reiese şi din denumirea lor *indicii dinamicii măsoară variaţia medie relativă a fenomenelor în timp*. Intervalul de timp cuprins între cele două

Scanned with OKEN Scanner

momente sau perioade de comparat trebuie astfel ales, încât fenomenul cercetat să nu fi suferit modificări calitative importante, să fie într-adevăr comparabil cu nivelul atins în perioada de bază.

25.3.1. Indicii simpli sau elementari

În vederea stabilirii formulelor de calcul al indicilor simpli, se consideră cantităţile şi preţurile a n produse într-o perioadă de $k+1$ ani (tabelul 25.1).

Tabelul 25.1. Cantităţile şi preţurile unitare pentru n produse pe ani

Anii	Cantităţile produsului i:							Preţurile produsului i:						
	1	2	3	...	i	...	n	1	2	3	...	i	...	n
0	q_{01}	q_{02}	q_{03}	⋯	q_{0i}	⋯	q_{0n}	p_{01}	p_{02}	p_{03}	⋯	p_{0i}	⋯	p_{0n}
1	q_{11}	q_{12}	q_{13}	⋯	q_{1i}	⋯	q_{1n}	p_{11}	p_{12}	p_{13}	⋯	p_{1i}	⋯	p_{1n}
2	q_{21}	q_{22}	q_{23}	⋯	q_{2i}	⋯	q_{2n}	p_{21}	p_{22}	p_{23}	⋯	p_{2i}	⋯	p_{2n}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
t	q_{t1}	q_{t2}	q_{t3}	⋯	q_{ti}	⋯	q_{tn}	p_{t1}	p_{t2}	p_{t3}	⋯	p_{ti}	⋯	p_{tn}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
k	q_{k1}	q_{k2}	q_{k3}	⋯	q_{ki}	⋯	q_{kn}	p_{k1}	p_{k2}	p_{k3}	⋯	p_{ki}	⋯	p_{kn}

Din tabelul 25.1 rezultă că q_{ti} şi p_{ti} reprezintă cantitatea şi respectiv, preţul produsului i în anul t.

Valoarea produsului i în anul t va fi egală cu produsul $q_{ti}\,p_{ti}$, adică se înmulţeşte q_{ti} din prima parte a tabelului 25.1. cu p_{ti} din partea a doua.

Pentru cunoaşterea variaţiei în timp a volumului fizic al fiecărui produs sau a dinamicii volumului fizic, se calculează indicii individuali ai volumului fizic. In funcţie de modul de alegere a bazei de comparaţie, indicii individuali pot fi de două feluri:

- *cu bază fixă*, notaţi cu $i^{q}_{t/0}$;

Scanned with OKEN Scanner

• *cu bază mobilă (sau în lanţ)*, notaţi cu $i^{q}_{t/t-1}$.

Indicii individuali cu bază fixă ai volumului fizic se calculează prin raportarea volumului fizic al produsului *i* din anul *t*, la volumul fizic al aceluiaşi produs din anul de bază. Dacă se consideră ca an de bază anul *0*, atunci se obţin următorii indici individuali ai volumului fizic şi ai preţurilor în perioada *t* faţă de *0*:

$$i^{q}_{t/0} = \frac{q_{ti}}{q_{0i}} \cdot 100 \quad si \quad i^{p}_{t/0} = \frac{p_{ti}}{p_{0i}} \cdot 100 \qquad \begin{matrix} t = \overline{0,k} \\ i = \overline{1,n} \end{matrix} \qquad (25.1)$$

Indicii individuali cu bază mobilă se obţin prin schimbarea bazei de la o perioadă la alta, astfel:

$$i^{q}_{t/t-1} = \frac{q_{ti}}{q_{t-1,i}} \cdot 100 \quad si \quad i^{p}_{t/t-1} = \frac{p_{ti}}{p_{t-1,i}} \cdot 100 \qquad \begin{matrix} t = \overline{0,k} \\ i = \overline{1,n} \end{matrix} \qquad (25.2)$$

Între indicii individuali cu bază fixă şi indicii individuali cu bază mobilă există o relaţie bine determinată şi anume, *produsul indicilor cu bază mobilă este egal cu indicele cu bază fixă al întregii perioade*:

$$\prod_{t=0}^{k} i^{q}_{t/t-1} = i^{q}_{k/0} \quad si \quad \prod_{t=0}^{k} i^{p}_{t/t-1} = i^{p}_{k/0} \qquad \begin{matrix} t = \overline{0,k} \\ i = \overline{1,n} \end{matrix} \qquad (25.3)$$

Din relaţia (25.3), rezultă că *raportul dintre doi indici cu bază fixă este egal cu indicele cu bază mobilă corespunzător perioadelor de raportare.*

$$i^{q}_{t/0} : i^{q}_{t-1/0} = i^{q}_{t/t-1} \quad si \quad i^{p}_{t/0} : i^{p}_{t-1/0} = i^{p}_{t/t-1} \qquad \begin{matrix} t = \overline{0,k} \\ i = \overline{1,n} \end{matrix} \qquad (25.4)$$

Relaţiile (25.3) şi (25.4) definesc tocmai una dintre proprietăţile cele mai importante ale indicilor individuali şi anume, cea de *circularitate* care permite trecerea de la indicii cu bază fixă la indicii cu bază mobilă şi invers.

O altă proprietate a indicilor individuali este cea de *reversibilitate*, care se poate scrie astfel:

Scanned with OKEN Scanner

$$i^{x}_{t/0} = \frac{1}{i^{x}_{0/t}} \quad sau \quad i^{x}_{t/0} \cdot i^{x}_{0/t} = 1 \tag{25.5}$$

În mod analog se vor obţine indicii individuali ai valorii unui produs în două perioade diferite.

$$i^{v}_{t/0} = \frac{v_{ti}}{v_{0i}} \cdot 100 = \frac{q_{ti}\ p_{ti}}{q_{0i}\ p_{0i}} \cdot 100 \qquad \begin{matrix} t=\overline{0,k} \\ t=\overline{1,n} \end{matrix} \tag{25.6}$$

deci

$$i^{v_i}_{t/0} = \frac{q_{ti}}{q_{0i}} \cdot \frac{p_{ti}}{p_{0i}} = i^{q_i}_{t/0} \cdot i^{p_i}_{t/0} \tag{25.7}$$

- indicii cu bază mobilă ai valorii produsului *i*:

$$i^{v_i}_{t/t-1} = \frac{v_{t,i}}{v_{t-1,i}} \cdot 100 = \frac{q_{ti}\ p_{ti}}{q_{t-1,i}\ p_{t-1,i}} \cdot 100 \qquad \begin{matrix} t=\overline{0,k} \\ i=\overline{1,n} \end{matrix} \tag{25.8}$$

$$i^{v_i}_{t/t-1} = \frac{q_{ti}}{q_{t-1,i}} \cdot \frac{p_{ti}}{p_{t-1,i}} = i^{q_i}_{t/t-1} \cdot i^{p_i}_{t/t-1} \tag{25.9}$$

Din relaţiile (25.7) şi (25.9), se observă că, indiferent de baza de comparaţie aleasă, *un indice individual al valorii unui produs este egal cu produsul dintre indicele individual al volumului fizic şi indicele individual al preţului corespunzător.*

Valoarea oricărui indice individual poate să fie mai mică, egală sau mai mare decât unitatea sau decât 100, atunci când indicele se exprimă în procente. În general *un indice arată de câte ori a crescut sau a scăzut nivelul unui fenomen în perioada de comparat faţă de nivelul aceluiaşi fenomen din perioada considerată de bază*, sau din aceeaşi perioadă de timp, dar situat în spaţii diferite. Dacă indicele este mai mic decât unitatea, înseamnă că nivelul unui fenomen a scăzut în perioada comparată faţă de perioada de bază; dacă valoarea indicelui este egală cu unitatea, înseamnă că fenomenul considerat nu a suferit nici o modificare (în sens de creştere sau scădere)

Scanned with OKEN Scanner

în perioada cercetată; iar dacă valoarea unui indice este supraunitară, aceasta arată că fenomenul studiat a crescut în perioada respectivă.

Atunci când indicii se exprimă în procente, perioada de bază este echivalată cu 100%, iar valoarea indicelui arată *cu cât a crescut* sau a scăzut, în procente, nivelul unui anumit fenomen în perioada de comparat faţă de perioada aleasă ca bază de comparaţie.

În cazul indicilor individuali, măsura variaţiei fenomenelor în perioada raportată faţă de perioada de bază se poate determina şi sub forma diferenţei dintre numărătorul şi numitorul indicelui respectiv, astfel:

- pentru preţurile unitare pe produs:

$$\Delta^{p}_{t/0} = p_{ti} - p_{oi} \qquad unde \qquad \Delta^{p}_{t/0} \lessgtr 0$$

şi

$$\Delta^{p}_{t/t-1} = p_{ti} - p_{t-1,i} \qquad (25.10)$$

- pentru volumul fizic al fiecărui produs:

$$\Delta^{q}_{t/0} = q_{ti} - q_{oi} \qquad unde \qquad \Delta^{q}_{t/0} \lessgtr 0$$

şi

$$\Delta^{q}_{t/t-1} = q_{ti} - q_{t-1,i} \qquad (25.11)$$

- pentru volumul valoric al fiecărui produs:

$$\Delta^{v}_{t/0} = v_{ti} - v_{oi} = q_{ti}p_{ti} - q_{0i}p_{0i} \qquad unde \qquad \Delta^{v}_{t/0} \lessgtr 0$$

şi

$$\Delta^{v}_{t/t-1} = v_{ti} - v_{t-1,i} = q_{ti}p_{ti} - q_{t-1,i}p_{t-1,i} \qquad (25.12)$$

Valorile pozitive ale acestor diferenţe exprimă creşteri ale fenomenelor individuale, cele negative corespund unor scăderi ale acestora, iar valoarea zero exprimă faptul că, în perioada luată în calcul, fenomenele respective nu au suferit nici o modificare, au rămas constante.

Scanned with OKEN Scanner

25.3.2. Indicii de grup sau sintetici

Cunoaşterea variaţiei în timp sau în spaţiu a fenomenelor individuale este importantă, dar nu este suficientă pentru formarea unei imagini de ansamblu, generale, asupra variaţiei fenomenelor colective de masă. Dinamica şi variaţia în spaţiu a fenomenelor colective se măsoară în statistică cu ajutorul *indicilor de grup* sau *sintetici*.

Indicii de grup se notează cu litere majuscule şi pot fi cu bază fixă şi cu bază mobilă.

25.3.2.1. Indicii agregaţi

Pentru construcţia indicilor agregaţi să considerăm din nou elementele din tabelul 25.1, şi pe baza lor, să calculăm volumul valoric al celor n produse în perioada 0 - k. Rezultatele se prezintă în tabelul 25.2. Pe ultima coloană a acestui tabel se află volumul valoric total al celor n produse din fiecare an.

În funcţie de modul de alegere a bazei de comparaţie se disting două categorii de indici agregaţi:

- cu bază fixă - notaţi cu $I_{t/0}$;
- cu bază mobilă - notaţi cu $I_{t/t-1}$.

Atât indicii cu bază fixă, cât şi cei cu bază mobilă, în funcţie de *modul de alegere a ponderilor*, pot fi *cu ponderi constante* sau *cu ponderi variabile*.

Indicii cu bază fixă şi ponderi constante ai volumului fizic şi ai preţurilor se calculează cu ajutorul formulelor:

$$I^{q}_{t/0} = \frac{\sum_{i=1}^{n} q_{ti} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} \qquad (25.13)$$

respectiv

$$I^{p}_{t/0} = \frac{\sum_{i=1}^{n} p_{ti} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} \qquad (25.14)$$

Scanned with OKEN Scanner

În ambele formule s-au utilizat ca ponderi nivelurile preţurilor şi respectiv, ale volumului fizic din anul de bază. În felul acesta, sumele de la numitorul formulelor (25.13) şi (25.14) reprezintă *valoarea reală* a produselor din perioada de bază, exprimată în preţurile aceleiaşi perioade, în timp ce numărătorii $\sum_{i=1}^{n} q_{ti} p_{0i}$ şi $\sum_{i=1}^{n} p_{ti} q_{0i}$ reprezintă *valori convenţionale* ale produselor din perioada de bază şi ale celor din perioada curentă.

Pe baza indicilor de grup în formă agregat, se poate determina influenţa variaţiei factorilor luaţi în studiu în perioada t faţă de perioada de bază, în mărimi absolute sub forma diferenţei dintre numărătorul şi numitorul formulei indicelui respectiv.

Tabelul 25.2. Valoarea produselor

Produsul / Anii	1	2	...	i	...	n	Total
0	$q_{01}p_{01}$	$q_{02}p_{02}$	.....	$q_{0i}p_{0i}$	.....	$q_{0n}p_{0n}$	$\sum_{i=0}^{n} q_{0i}p_{0i}$
1	$q_{11}p_{11}$	$q_{12}p_{12}$	.....	$q_{1i}p_{1i}$	.....	$q_{1n}p_{1n}$	$\sum_{i=0}^{n} q_{1i}p_{1i}$
2	$q_{21}p_{21}$	$q_{22}p_{22}$	.....	$q_{2i}p_{2i}$	.....	$q_{2n}p_{2n}$	$\sum_{i=0}^{n} q_{2i}p_{2i}$
.....	.....	.....	.....	.....	.....	.....	.....
t	$q_{t1}p_{t1}$	$q_{t2}p_{t2}$	.....	$q_{ti}p_{ti}$	.....	$q_{tn}p_{tn}$	$\sum_{i=0}^{n} q_{ti}p_{ti}$
.....	.....	.....	.....	.....	.....	.....	.....
k	$q_{k1}p_{k1}$	$q_{k2}p_{k2}$	.....	$q_{ki}p_{ki}$	.....	$q_{kn}p_{kn}$	$\sum_{i=0}^{n} q_{ki}p_{ki}$

Scanned with OKEN Scanner

Astfel, în cazul indicelui de grup al volumului fizic (25.13), rezultă:

$$\Delta^{q}_{t/0} = \sum_{i=1}^{n} q_{ti} p_{0i} - \sum_{i=1}^{n} q_{0i} p_{0i} \qquad (25.15)$$

Acest indicator reprezintă diferenţa dintre suma valorii tuturor celor n produse în anul t exprimate în preţurile perioadei de bază 0 şi valoarea aceloraşi n produse din perioada de bază exprimate în preţurile perioadei de bază. Din formula (25.14) rezultă:

$$\Delta^{p}_{t/0} = \sum_{i=1}^{n} p_{ti} q_{0i} - \sum_{i=1}^{n} p_{0i} q_{0i} \qquad (25.16)$$

Un indicator important de caracterizare a nivelului de trai al populaţiei îl constituie *indicele costului vieţii*. Acest indice măsoară variaţia medie relativă a preţurilor mărfurilor şi tarifelor serviciilor care formează consumul.

Indicii agregaţi cu bază fixă şi cu ponderi variabile ai volumului fizic şi ai preţurilor se calculează astfel:

$$I^{q}_{t/0} = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{ti}} \qquad (25.17)$$

respectiv

$$I^{p}_{t/0} = \frac{\sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{ti}} \qquad (25.18)$$

Indicele care măsoară în mod concomitent variaţia preţurilor şi a volumului fizic al tuturor produselor este *indicele agregat al valorii*, definit de formula:

Scanned with OKEN Scanner

$$I^{v}_{t/0} = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{0i}} \qquad (25.19)$$

şi

$$I^{v}_{t/t-1} = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{t-1,i} p_{t-1,i}} \qquad (25.20)$$

Deoarece indicele agregat al valorii este un indice bifactorial, la fel ca şi indicele individual al valorii, se poate descompune în factori. Între indicele valoric, indicele volumului fizic şi indicele preţurilor există relaţia:

$$I^{v}_{t/0} = I^{q}_{t/0} \cdot I^{p}_{t/0} \quad \text{şi} \quad I^{v}_{t/t-1} = I^{q}_{t/t-1} \cdot I^{p}_{t/t-1} \qquad (25.21)$$

adică

$$\frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{0i}} = \frac{\sum_{i=1}^{n} q_{ti} p_{oi}}{\sum_{i=1}^{n} q_{0i} p_{0i}} \cdot \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{ti} p_{0i}} \qquad (25.22)$$

respectiv

$$\frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{t-1,i} p_{t-1,i}} = \frac{\sum_{i=1}^{n} q_{ti} p_{t-1,i}}{\sum_{i=1}^{n} q_{t-1,i} p_{t-1,i}} \cdot \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{ti} p_{t-1,i}} \qquad (25.23)$$

În cazul influenţelor factorilor exprimate în mărimi absolute, se va verifică relaţia:

$$\Delta^{v}_{t/0} = \Delta^{q}_{t/0} + \Delta^{p}_{t/0} \qquad (25.24)$$

Scanned with OKEN Scanner

respectiv

$$\Delta^{v}_{t/t-1} = \Delta^{q}_{t/t-1} + \Delta^{p}_{t/t-1} \tag{25.25}$$

Precizăm faptul că relaţiile (25.21) şi (25.24) se verifică atât în cazul indicilor cu bază fixă (25.22), cât şi în cazul indicilor cu bază mobilă (25.23), dacă unul dintre indicii factoriali simpli este ponderat cu nivelul factorului de ponderare din perioada de bază şi celălalt, cu nivelul corespunzător din perioada curentă.

La rândul lor, indicii sub formă agregat cu bază mobilă pot fi construiţi cu ponderi constante sau cu ponderi variabile, astfel:

- în cazul utilizării unei ponderi constante:

$$I^{q}_{t/t-1} = \frac{\sum_{i=1}^{n} q_{ti}\, p_{0i}}{\sum_{i=1}^{n} q_{t-1,i}\, p_{0i}} \tag{25.26}$$

şi

$$I^{p}_{t/t-1} = \frac{\sum_{i=1}^{n} p_{ti}\, q_{0i}}{\sum_{i=1}^{n} p_{t-1,i}\, q_{0i}} \tag{25.27}$$

- în cazul utilizării unor ponderi variabile:

$$I^{q}_{t/t-1} = \frac{\sum_{i=1}^{n} q_{ti}\, p_{t-1,i}}{\sum_{i=1}^{n} q_{t-1,i}\, p_{t-1,i}} \tag{25.28}$$

şi

Scanned with OKEN Scanner

$$I^{p}_{t/t-1} = \frac{\sum_{i=1}^{n} p_{ti} q_{t-1,i}}{\sum_{i=1}^{n} p_{t-1,i} q_{t-1,i}} \qquad (25.29)$$

- în cazul reflectării concomitente a influenţei ambilor factori:

$$I^{v}_{t/t-1} = \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{t-1,i} p_{t-1,i}} \qquad (25.30)$$

Statisticianul german *Etienne Laspeyres (1834-1913)* a propus ca indicii agregaţi să se construiască cu ajutorul ponderilor din perioada de bază (25.13) şi (25.14). Aceşti indici se utilizează în studierea dinamicii volumului fizic al produselor, încercând a se elimina influenţa factorului de comensurare preţ.

Formulele indicilor agregat (25.17) şi (25.18), în care se utilizează ca ponderi elemente ale perioadei curente, poartă denumirea de *indici de tip Paasche*, după numele economistului german *Herman Paasche (1851-1925)*, care a propus pentru prima dată acest sistem de ponderare. Indicii de tip Paasche se aplică în practica statistică pentru calculul dinamicii preţurilor, întrucât interesează în ce măsură au variat preţurile produselor din perioada curentă.

Din punct de vedere metodologic, ambele formule sunt la fel de satisfăcătoare, prezentând avantaje şi dezavantaje asemănătoare.

Indicele de tip Paasche şi indicele de tip Laspeyres, calculaţi pentru acelaşi fenomen vor avea valori diferite în măsura în care ponderile din perioada de bază diferă de cele din perioada curentă.

Diferenţa dintre un indice de tip Paasche şi un indice de tip Laspeyres a fost determinată de către statisticianul german *Ladislau von Bortkiewicz (1868-1931)*. Relaţia lui *Bortkiewicz* pentru indicii volumului fizic are următoarea expresie:

Scanned with OKEN Scanner

$$\frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{ti}} : \frac{\sum_{i=1}^{n} q_{ti} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} = 1 + r_{i^q_{t/0}, i^p_{t/0}} \, V_{i^q_{t/0}} \cdot V_{i^p_{t/0}} \qquad (25.31)$$

unde

$$r_{i^q_{t/0}, i^p_{t/0}} = \frac{\sum_{i=1}^{n} (i^q_{t/0} - I^q_{t/0})(i^p_{t/0} - I^p_{t/0}) q_{0i} p_{0i}}{\sigma_{i^q_{t/0}} \sigma_{i^p_{t/0}} \sum_{i=1}^{n} q_{0i} p_{0i}}$$

$$\sigma_{i^q_{t/0}} = \frac{\sum_{i=1}^{n} (i^q_{t/0} - I^p_{t/0})^2 q_{0i} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}}$$

$$\sigma_{i^p_{t/0}} = \frac{\sum_{i=1}^{n} (i^p_{t/0} - I^p_{t/0})^2 q_{0i} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}}$$

Statisticianul englez *Francis Ysidor Edgeworth (1845-1926)* a propus ca indicele agregat să fie ponderat cu media aritmetică simplă a ponderilor din perioadele de comparat, astfel:

- în cazul volumului fizic:

$$I^q_{t/0} = \frac{\sum_{i=1}^{n} q_{ti} \frac{p_{0i} + p_{ti}}{2}}{\sum_{i=1}^{n} q_{0i} \frac{p_{0i} + p_{ti}}{2}} = \frac{\sum_{i=1}^{n} q_{ti} (p_{0i} + p_{ti})}{\sum_{i=1}^{n} q_{0i} (p_{0i} + p_{ti})} = \frac{\sum_{i=1}^{n} q_{ti} p_{0i} + \sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{0i} + \sum_{i=1}^{n} q_{0i} p_{ti}} \qquad (25.32)$$

Scanned with OKEN Scanner

- în cazul preţurilor:

$$I^{p}_{t/0} = \frac{\sum_{i=1}^{n} p_{ti} \frac{q_{0i} + q_{ti}}{2}}{\sum_{i=1}^{n} p_{0i} \frac{q_{0i} + q_{0i}}{2}} = \frac{\sum_{i=1}^{n} p_{ti} q_{0i} + \sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{0i} + \sum_{i=1}^{n} p_{0i} q_{ti}} \qquad (25.33)$$

O altă soluţie în problema ponderării indicilor agregat a fost propusă de statisticianul *S. Drobisch în anul 1871. Drobisch* consideră că este necesar să se ia în calculul indicelui agregat atât ponderile din perioada curentă, cât şi cele din perioada de bază, iar indicele agregat să se calculeze ca o medie aritmetică simplă a indicilor de tip *Laspeyres* şi a celor de tip *Paasche* după formula:

- în cazul volumului fizic:

$$I^{q}_{t/0} = \frac{\frac{\sum_{i=1}^{n} q_{ti} p_{0i}}{\sum_{i=1}^{n} q_{0i} p_{0i}} + \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{ti}}}{2} \qquad (25.34)$$

- în cazul preţurilor:

$$I^{p}_{t/0} = \frac{\frac{\sum_{i=1}^{n} p_{ti} q_{0i}}{\sum_{i=1}^{n} p_{0i} q_{0i}} + \frac{\sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{ti}}}{2} \qquad (25.35)$$

Pentru exemplificarea calculului şi analizei indicilor individuali şi a celor de grup sub formă agregat, să considerăm datele convenţionale din tabelul 25.3.

Variaţia relativă a volumului fizic al fiecărui produs în perioada curentă faţă de perioada de bază este dată de indicii individuali, calculaţi după formula (25.1), astfel:

- Pentru produsul A :

Scanned with OKEN Scanner

$$i^{q}_{t/0} = \frac{q_t}{q_0} = \frac{300}{200} = 1{,}5$$

sau $$i^{q}_{t/0} \cdot 100 - 100 = 150 - 100 = 50\%$$

iar $$\Delta^{q}_{t/0} = q_t - q_0 = 300 - 200 = 100\ kg$$

Deci, volumul fizic al produsului A a crescut în perioada curentă, faţă de perioada de bază, cu 100 kg sau de 1,5 ori sau, în procente, cu 50%.

- Pentru produsul B:

$$i^{q}_{t/0} = \frac{300}{300} = 1 \qquad \text{sau} \qquad \cdot i^{q}_{t/0} \cdot 100 - 100 = 0$$

iar

$$\Delta^{q}_{t/0} = 300 - 300 = 0$$

Aceasta înseamnă că volumul fizic al produsului *B* a rămas nemodificat în perioada curentă faţă de perioada de bază.

Tabelul 25.3

Produsul	U.M.	Perioada de bază		Perioada curentă	
		Volumul fizic	Preţurile unitare (lei)	Volumul fizic	Preţurile unitare (lei)
A	kg	200	40	300	40
B	m	300	20	300	15
C	buc	500	30	450	35

- Pentru produsul *C*:

$$i^{q_i}_{t/0} = \frac{450}{500} = 0{,}9 \qquad sau \qquad i^{q_i}_{t/0} \cdot 100 - 100 = -10\%$$

Scanned with OKEN Scanner

iar $\Delta^{q_i}_{t/0} = 450 - 500 = -50 \ buc$

Volumul fizic al produsului C s-a micşorat în perioada curentă, faţă de perioada de bază, cu 50 bucăţi sau de 0,9 ori sau în procente cu 10%.

Calculele elementelor intermediare necesare pentru determinarea indicilor de grup se prezintă sistematizat în tabelul 25.4.

Tabelul 25.4.

Produsul i	q_{0i}	p_{0i}	q_{ti}	p_{ti}	$q_{0i}p_{0i}$	$q_{ti}p_{0i}$	$q_{ti}p_{ti}$	$q_{0i}p_{ti}$
A	200	40	300	40	8000	12000	12000	8000
B	300	20	300	15	6000	6000	4500	4500
C	500	30	450	35	15000	13500	15750	17500
Total	-	-	-	-	29000	31500	32250	30000

Variaţia volumului fizic al tuturor produselor în perioada curentă, faţă de perioada de bază , se poate determina cu ajutorul indicelui agregat al volumului fizic de tip *Laspeyres*, dat de formula (25.13) pentru $t = \overline{0,k}$ şi $i = 1,2,3$ astfel:

$$I^{q}_{t/0} = \frac{\sum_{i=1}^{3} q_{ti}p_{0i}}{\sum_{i=1}^{3} q_{0i}p_{0i}} = \frac{31500}{29000} = 1,0862 \quad sau \quad I^{q}_{t/0} \cdot 100 - 100 = 8,62\%$$

iar

$$\Delta^{q}_{t/0} = \sum_{i=1}^{3} q_{tj}p_{0j} - \sum_{i=1}^{3} q_{0i}p_{0i} = 31500 - 29000 = 2500 \quad lei$$

Rezultă că volumul fizic al celor trei produse (exprimate valoric) a crescut în perioada curentă , faţă de perioada de bază, cu 2500 lei sau de 1,0862 ori sau cu 8,62%.

Scanned with OKEN Scanner

Această creştere a volumului fizic, exprimat valoric pe total produse se datorează, pe de o parte, faptului că volumul fizic al produsului *A* a crescut mai mult decât a scăzut cel al produsului *C* iar, pe de altă parte, în mod implicit şi într-o măsură mai mică , factorului de ponderare preţ,.

Dinamica preţurilor pe fiecare produs şi pe total se determină cu ajutorul indicilor individuali (25.1) şi a celor agregat de tip *Paasche* (25.18):

- pentru produsul *A*:

$$i_{t/0}^{p_i} = \frac{p_t}{p_0} = \frac{40}{40} = 1 \quad sau \quad i_{t/0}^{p_i} \cdot 100 - 100 = 0$$

iar $$\Delta_{t/0}^{p_i} = p_t - p_0 = 40 - 40 = 0$$

- pentru produsul *B*:

$$i_{t/0}^{p_i} = \frac{15}{20} = 0{,}75 \quad sau \quad i_{t/0}^{p_i} \cdot 100 - 100 = -25\%$$

iar $$\Delta_{t/0}^{p_i} = 15 - 20 = -5 \quad lei$$

- pentru produsul *C*:

$$i_{t/0}^{p_i} = \frac{35}{30} = 1{,}1667 \quad sau \quad i_{t/0}^{p_i} \cdot 100 - 100 = 16{,}67\%$$

iar $$\Delta_{t/0}^{p_i} = 35 - 30 = 5 \quad lei$$

- pe total produse *A*, *B* şi *C*:

$$I_{t/0}^{p} = \frac{\sum_{i=1}^{3} p_{ti} q_{ti}}{\sum_{i=1}^{3} p_{0i} q_{ti}} = \frac{32250}{31500} = 1{,}0238 \quad sau \quad I_{t/0}^{p} \cdot 100 - 100 = 2{,}38\%$$

iar $$\Delta_{t/0}^{p} = \sum_{i=1}^{3} p_{ti} q_{ti} - \sum_{i=1}^{3} p_{0i} q_{ti} = 32250 - 31500 = 750 \quad lei$$

Scanned with OKEN Scanner

Rezultă că variaţia diferită a preţurilor celor trei produse în perioada curentă , faţă de perioada de bază, a determinat o creştere a preţurilor pe ansamblu de 1,0238 ori sau cu 2,38%. Această creştere a preţurilor a dus la o creştere a volumului valoric al producţiei cu 750 lei.

Ca şi în cazul indicelui volumului fizic (în general pentru toţi indicii sub formă agregat), valoarea indicelui agregat al preţurilor este influenţată şi de factorul de ponderare q_{ti}.

Variaţia volumului valoric al celor trei produse, pe fiecare produs şi pe total se determină cu ajutorul indicilor individuali ai valorii (25.6) şi a celui agregat (25.19), astfel:

• pentru produsul *A*:

$$i_{t/0}^{v_i} = \frac{q_{ti}p_{ti}}{q_{0i}p_{oi}} = \frac{12000}{8000} = 1{,}5 \quad sau \quad i_{t/0}^{v_i} \cdot 100 - 100 = 50\%$$

iar $$\Delta_{t/0}^{vi} = q_{ti}p_{ti} - q_{0i}p_{oi} = 12000 - 8000 = 4000 \quad lei$$

Pe baza relaţiei (25.7) rezultă:

$$i_{t/0}^{v_i} = i_{t/0}^{q} \cdot i_{t/0}^{p} = 1.5 \cdot 1 = 1{,}5$$

Valoarea produsului *A* a crescut în perioada curentă faţă de perioada de bază cu 50%, ca urmare a creşterii volumului fizic al acestui produs cu acelaşi procent şi a menţinerii constante a preţului.

• pentru produsul *B*:

$$i_{t/0}^{v_i} = \frac{4500}{6000} = 0{,}75 \quad sau \quad i_{t/0}^{v_i} \cdot 100 - 100 = -25\%$$

iar $$\Delta_{t/0}^{v_i} = 4500 - 6000 = -1500 \quad lei$$

dar $$i_{t/0}^{v_i} = 1 \cdot 0{,}75 = 0{,}75$$

Aceasta înseamnă că valoarea produsului *B* a scăzut în perioada curentă faţă de perioada de bază cu 25%, ca urmare a reducerii cu 25% a preţului acestui produs.

• pentru produsul *C*:

Scanned with OKEN Scanner

$$i_{t/0}^{vi} = \frac{15750}{15000} = 1{,}05 \qquad sau \qquad i_{t/0}^{v_i} \cdot 100 - 100 = 5\%$$

iar $$\Delta_{t/0}^{v_i} = 15750 - 15000 = 750 \quad lei$$

dar $$i_{t/0}^{v_i} = 0{,}9 \cdot 1{,}1667 = 1{,}05$$

Deci, valoarea produsului *C* a crescut în perioada curentă faţă de perioada de bază cu 5%, datorită creşterii cu 16,67% a preţurilor şi a reducerii cu 10% a volumului fizic al produsului luat în calcul.

• pe total produse:

$$I_{t/0}^{v} = \frac{\sum_{i=1}^{3} q_{ti} p_{ti}}{\sum_{i=1}^{3} q_{0i} p_{0i}} = \frac{32250}{29000} = 1{,}112 \quad \text{sau} \quad I_{t/0}^{v} \cdot 100 - 100 = 11{,}2\%$$

iar $$\Delta_{t/0}^{v} = \sum_{i=1}^{3} q_{ti} p_{ti} - \sum_{i=1}^{3} q_{0i} p_{0i} = 32250 - 29000 = 3250 \quad \text{lei}$$

Rezultă că valoarea produselor *A*, *B* şi *C* a crescut în perioada curentă faţă de perioada de bază cu 11,2% sau, în valoare absolută, cu 3250 lei.

Indicele valoric sub formă agregat mai poate fi determinat şi cu ajutorul relaţiei (25.21), astfel:

$$I_{t/0}^{v} = I_{t/0}^{q} \cdot I_{t/0}^{p} = 1{,}0862 \cdot 1{,}0238 = 1{,}112$$

iar în mărimi absolute:

$$\Delta_{t/0}^{v} = \Delta_{t/0}^{q} + \Delta_{t/0}^{p} = 2500 + 750 = 3250 \quad \text{lei}$$

Indicele agregat al volumului fizic şi cel al preţurilor se pot calcula şi cu ajutorul formulelor stabilite de *Edgeworth* (25.32) şi (25.33), cele ale lui *Drobisch* (25.34) şi (25.35), pe baza tabelului 25.5.

Scanned with OKEN Scanner

Tabelul 25.5

Produsele i	q_{0i}	p_{0i}	q_{ti}	p_{ti}	p'_i	$q_{oi}p'_i$	$q_{ti}p'_i$	q'_i	q'_ip_{0i}	q'_ip_{ti}
A	200	40	300	40	80	16000	24000	500	20000	20000
B	300	20	300	15	35	10500	10500	600	12000	9000
C	500	30	450	35	65	32500	29250	950	28500	33250
Total	-	-	-	-	-	59000	63750	-	60500	62250

Indicii de tip *Edgeworth*:

$$I^q_{t/0} = \frac{\sum_{i=1}^{3} q_{ti}(p_{0i} + p_{ti})}{\sum_{i=1}^{3} q_{0i}(p_{0i} + p_{ti})} = \frac{\sum_{i=1}^{3} q_{ti}p'_i}{\sum_{i=1}^{3} q_{0i}p'_i} = \frac{63750}{59000} = 1{,}0805$$

şi

$$I^p_{t/0} = \frac{\sum_{i=1}^{3} p_{ti}(q_{0i} + q_{ti})}{\sum_{i=1}^{3} p_{0i}(q_{0i} + q_{ti})} = \frac{\sum_{i=1}^{3} p_{ti}q'_i}{\sum_{i=1}^{3} p_{0i}q'_i} = \frac{62250}{60500} = 1{,}0289$$

Pentru a calcula indicele volumului fizic şi al preţurilor de tip *Drobisch*, va trebui să determinăm mai întâi indicele agregat de tip *Paasche* pentru volumul fizic şi indicele agregat de tip *Laspeyres* pentru preţuri, astfel:

$$I^q_{t/0} = \frac{\sum_{i=1}^{3} q_{ti}p_{ti}}{\sum_{i=1}^{3} q_{0i}p_{ti}} = \frac{32250}{30000} = 1{,}075$$

şi

Scanned with OKEN Scanner

$$I_{t/0}^{p} = \frac{\sum_{i=1}^{3} q_{0i} p_{ti}}{\sum_{i=1}^{3} q_{0i} p_{0i}} = \frac{30000}{29000} = 1{,}0344$$

Observăm că indicele volumului fizic şi cel al preţurilor iau valori diferite, atunci când se utilizează sisteme de ponderare diferite.

Indicii de tip *Drobisch*:

$$I_{t/0}^{q} = \left(\frac{\sum_{i=1}^{3} q_{ti} p_{0i}}{\sum_{i=1}^{3} q_{0i} p_{0i}} + \frac{\sum_{i=1}^{3} q_{ti} p_{ti}}{\sum_{i=1}^{3} q_{0i} p_{ti}} \right) :2 = (1{,}0862 + 1{,}075):2 = 1{,}0806$$

şi

$$I_{t/0}^{p} = \left(\frac{\sum_{i=1}^{3} q_{0i} p_{ti}}{\sum_{i=1}^{3} q_{0i} p_{0i}} + \frac{\sum_{i=1}^{3} q_{ti} p_{ti}}{\sum_{i=1}^{3} q_{ti} p_{0i}} \right) :2 = (1{,}0344 + 1{,}0238):2 = 1{,}0291$$

Se observă că indicii de tip *Edgeworth* şi cei de tip *Drobisch* au valori apropiate pentru aceleaşi fenomene, iar aceste valori se situează între valorile indicilor de tip *Laspeyres* şi a celor de tip *Paasche*.

Indicii sub formă de agregat se pot construi şi pentru mai mult de doi factori.

În cazul a trei factori x, y şi z, indicele agregat, care oglindeşte influenţa tuturor celor trei factori, se va calcula după formulele:

• cu bază fixă:

$$I_{t/0}^{xyz} = \frac{\sum_{i=0}^{n} x_{ti} y_{ti} z_{ti}}{\sum_{i=0}^{n} x_{0i} y_{0i} z_{0i}} \qquad \begin{matrix} t = \overline{0,k} \\ i = \overline{0,n} \end{matrix} \qquad (25.36)$$

Scanned with OKEN Scanner

şi

$$\Delta_{t/0}^{xyz} = \sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} - \sum_{i=0}^{n} x_{0i} y_{0i} z_{0i}$$

• cu bază mobilă:

$$I_{t/t-1}^{xyz} = \frac{\sum_{i=0}^{n} x_{ti} y_{ti} z_{ti}}{\sum_{i=0}^{n} x_{t-1,i} y_{t-1,i} z_{t-1,i}} \qquad (25.37)$$

iar

$$\Delta_{t/t-1}^{xyz} = \sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} - \sum_{i=0}^{n} x_{t-1,i} y_{t-1,i} z_{t-1,i}$$

În cazul luării în calcul a patru factori x, y, z, v indicii sub formă agregat se vor calcula după formulele:

• cu bază fixă:

$$I_{t/0}^{xyzv} = \frac{\sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} v_{ti}}{\sum_{i=0}^{n} x_{0i} y_{0i} z_{0i} v_{0i}} \qquad (25.38)$$

iar

$$\Delta_{t/0}^{xyzv} = \sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} v_{ti} - \sum_{i=0}^{n} x_{0i} y_{0i} z_{0i} v_{0i}$$

• cu bază mobilă:

$$I_{t/t-1}^{xyzv} = \frac{\sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} v_{ti}}{\sum_{i=0}^{n} x_{t-1,i} y_{t-1,i} z_{t-1,i} v_{t-1,i}} \qquad (25.39)$$

iar $$\Delta_{t/t-1}^{xyzv} = \sum_{i=0}^{n} x_{ti} y_{ti} z_{ti} v_{ti} - \sum_{i=0}^{n} x_{t-1,i} y_{t-1,i} z_{t-1,i} v_{t-1,i}$$

Scanned with OKEN Scanner

25.3.2.2. Indicele mediu aritmetic ponderat

O formă transformată a indicelui agregat o constituie *indicele mediu aritmetic ponderat, care se calculează ca o medie aritmetică a indicilor individuali ai fenomenului studiat ponderaţi cu elementele de la numitorul formulei indicelui agregat corespunzător.*

Astfel, atunci când se cunosc indicii individuali ai volumului fizic pentru un grup de produse şi valorile acestor produse din perioada de bază, dinamica volumului fizic al producţiei grupei de produse se poate calcula după formula:

$$I^q_{t/0} = \frac{\sum_{i=1}^{n} i^q_{t/0} \cdot q_{0i}p_{0i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} \qquad (25.40)$$

Dacă se substituie în formula (25.40) $i^q_{t/0} = \frac{q_{ti}}{q_{0i}}$, rezultă:

$$I^q_{t/0} = \frac{\sum_{i=1}^{n} \frac{q_{ti}}{q_{0i}} \cdot q_{0i}p_{0i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} = \frac{\sum_{i=1}^{n} q_{ti}p_{0i}}{\sum_{i=1}^{n} q_{0i}p_{0i}} \qquad (25.41)$$

Relaţia (25.41) arată că indicele mediu aritmetic ponderat este egal cu indicele agregat corespunzător.

Pentru exemplificarea calculului şi analizei acestui indice, considerăm că se cunosc indicii individuali ai volumului fizic pentru cele trei produse din exemplul anterior (tabelul 25.3) şi valorile celor trei produse din perioada de bază, exprimate în preţurile perioadei (tabelul 25.6).

Substituind datele din tabelul 25.6 în formula (25.40), se obţine:

$$I^q_{t/0} \quad \frac{31500}{29000} \quad 1{,}0862$$

Scanned with OKEN Scanner

Tabelul 25.6.

Produsele	$i^q_{t/0}$ 166	$q_{oi}p_{oi}$ 167	$i^q_{t/0}q_{oi}p_{oi}$
A	1,5	8000	12000
B	1,0	6000	6000
C	0,9	15000	13500
Total	-	29000	31500

Se observă că s-a obţinut aceeaşi valoare a indicelui de grup al volumului fizic ca şi în cazul utilizării formulei indicelui agregat de tip *Laspeyres*.

25.3.2.3. Indicele mediu armonic ponderat

Acest indice se utilizează în activitatea statistică atunci când se cunosc indicii individuali ai fenomenului cercetat şi elementele de la numărătorul formulei indicelui agregat corespunzător. *Indicele de grup mediu armonic ponderat se calculează ca o medie armonică ponderată a indicilor individuali ai fenomenului cercetat ponderaţi cu elementele de la numărătorul formulei indicelui agregat corespunzător.* Se aplică în practica statistică, de regulă pentru determinarea dinamicii preţurilor. Indicele mediu armonic ponderat al preţurilor se calculează pe baza formulei:

$$I^p_{t/0} = \frac{\sum_{i=1}^{n} p_{ti}q_{ti}}{\sum_{i=1}^{n} \frac{1}{i^p_{t/0}} p_{ti}q_{ti}} \qquad (25.42)$$

Dacă în formula (25.42) se substituie $i^p_{t/0} = \frac{p_{ti}}{p_{0i}}$, rezultă:

$$I^p_{t/0} = \frac{\sum_{i=1}^{n} p_{ti}q_{ti}}{\sum_{i=1}^{n} \frac{1}{\frac{p_{ti}}{p_{0i}}} p_{ti}q_{ti}} = \frac{\sum_{i=1}^{n} p_{ti}q_{ti}}{\sum_{i=1}^{n} p_{0i}q_{ti}} \qquad (25.43)$$

Din relaţia (25.43) rezultă că indicele mediu armonic este echivalent cu indicele agregat corespunzător.

Să considerăm, pentru exemplificare, că pentru cele trei produse din exemplul anterior, se cunosc indicii individuali ai preţurilor şi valoarea produselor din perioada curentă, exprimată în preţurile acestei perioade (tabelul 25.7).

Tabelul 25.7.

Produsul	$i_{t/0}^{p_i}$	$q_{ti}p_{ti}$	$\frac{q_{ti}p_{ti}}{i_{t/0}^{p_i}}$
A	1	12000	12000
B	0,75	4500	6000
C	1,1667	15750	13500
Total	-	32250	31500

25.3.2.4. Indicele mediu geometric

W.S.Jevons într-un studiu binecunoscut *A serious fall in the value of gold ascertained*, publicat în *1863*, calculează pentru prima dată indicii cu ajutorul formulei mediei geometrice simple din indici individuali, prin extragerea rădăcinii de ordinul *n* din produsul a *n* indici simpli, după formula:

$$I_{t/0}^{p} = \sqrt[n]{\frac{p_{t1}}{p_{01}} \cdot \frac{p_{t2}}{p_{02}} \cdots \frac{p_{ti}}{p_{0i}} \cdots \frac{p_{tn}}{p_{0n}}} \qquad (25.44)$$

Jevons a studiat problema scăderii preţului aurului, ca urmare a descoperirii de noi zăcăminte în *California* şi *Australia* în jurul anului *1850*. El a luat în calcul preţurile a *39* de produse din anii *1845-1862*, considerând ca preţuri de bază pentru fiecare produs, media aritmetică a preţurilor din perioada *1845-1850*. Apoi a împărţit preţurile, corespunzătoare pentru fiecare din cei *7* ani ai perioadei la nivelul lor mediu de bază, obţinând astfel indicii individuali ai preţurilor, pe baza cărora a determinat nivelul lor mediu cu ajutorul mediei geometrice simple.

Deficienţa principală a indicelui lui Jevons constă în faptul că nu a luat în considerare volumul fizic al produselor drept ponderе.

Cu toate acestea, metoda lui Jevons a stat mai târziu la baza faimoasei formule a *indicelui ideal al preţurilor* elaborată de către *C.M.Walsh* în lucrarea sa

Scanned with OKEN Scanner

Measurement of General Exchange Value, într-o notă de subsol al unei pagini în anul *1912*. Aceeaşi formulă este publicată de *A.C.Pigou* în anul *1912* în lucrarea *Wealth and Welfare* la pagina *46*, dar nu extrage rădăcina pătrată din produs. Această deficienţă este înlăturată de autor în anul *1920*, în lucrarea *Economics of Welfare*, la pagina *78*.

Problema construirii unui indice ideal este preluată şi analizată sub multiple aspecte de către statisticianul american *Irving Fisher*, care o prezintă într-o comunicare făcută la *American Statistical Association* în anul *1923*.

Indicele mediu geometric îmbină calităţile indicelui agregat de tip Laspeyres cu cele ale indicelui de tip Paasche şi se calculează ca o medie geometrică simplă a celor două tipuri fundamentale ale indicelui agregat unifactorial, după formulele:

- în cazul volumului fizic:

$$I_{t/0}^{q} = \sqrt{\frac{\sum_{i=1}^{n} q_{ti} p_{oi}}{\sum_{i=1}^{n} q_{0i} p_{oi}} \cdot \frac{\sum_{i=1}^{n} q_{ti} p_{ti}}{\sum_{i=1}^{n} q_{0i} p_{ti}}} \qquad (25.45)$$

- în cazul preţurilor:

$$I_{t/0}^{p} = \sqrt{\frac{\sum_{i=1}^{n} p_{ti} q_{oi}}{\sum_{i=1}^{n} p_{0i} q_{oi}} \cdot \frac{\sum_{i=1}^{n} p_{ti} q_{ti}}{\sum_{i=1}^{n} p_{0i} q_{ti}}} \qquad (25.46)$$

Valoarea indicelui mediu geometric ponderat al preţurilor, în mod logic, este cuprinsă între valoarea indicelui agregat al preţurilor de tip Laspeyres şi valoarea indicelui agregat al preţurilor de tip Paasche, a căror medie este.

Pentru exemplificarea calculului şi analizei indicelui mediu geometric, să considerăm datele din tabelul 25.4.

Indicele mediu geometric al preţurilor celor trei produse se obţine pe baza fomulei (25.46), astfel:

$$I_{t/0}^{p} = \sqrt{\frac{\sum_{i=1}^{3} p_{ti} q_{oi}}{\sum_{i=1}^{3} p_{0i} q_{oi}} \cdot \frac{\sum_{i=1}^{3} p_{ti} q_{ti}}{\sum_{i=1}^{3} p_{0i} q_{ti}}} = \sqrt{\frac{30000}{29000} \cdot \frac{32250}{31500}} = \sqrt{1{,}0344 \cdot 1{,}0238} \cong 1{,}029$$

Scanned with OKEN Scanner

25.3.2.5. Indici calculaţi ca raport între două medii aritmetice ponderate

Sistemul indicilor calculaţi ca raport între două medii aritmetice ponderate este compus din trei categorii de indici:

- *indici bifactoriali*;
- *indici unifactoriali cu structură fixă*;
- *indici unifactoriali cu variaţii ale structurii.*

Indicii bifactoriali reflectă influenţa concomitentă a doi factori şi anume, variaţia caracteristicii x şi a structurii acesteia. Deoarece în calculul nivelului mediu intră doi factori - nivelurile individuale ale fenomenului studiat x_i şi structura colectivităţii $\left(\frac{f_{ti}}{\sum_{i=1}^{n} f_{ti}}\right)$, indicele bifactorial va arăta influenţa concomitentă a celor doi factori şi se va calcula după formula:

$$I_{t/o}^{xf} = \frac{\bar{x}_t}{\bar{x}_0} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{0i}}{\sum_{i=1}^{n} f_{0i}} \qquad (25.47)$$

Indicele bifactorial se poate descompune în doi indici unifactoriali : unul care să arate numai variaţia caracteristicii (*indicele cu structură fixă*) şi altul care să arate numai influenţa variaţiei structurii colectivităţii (*indicele variaţiei structurii*).

Substituind în formula (25.47) elementul de structură din perioada de bază f_{0i} cu cel din perioada curentă f_{ti}, se obţine formula uzuală a *indicelui cu structură fixă*:

$$I_{t/o}^{x} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{ti}} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} x_{0i} f_{ti}} \qquad (25.48)$$

Scanned with OKEN Scanner

Din relaţia (25.48), se observă că indicele cu structură fixă are forma unui indice agregat de tip *Paasche*, cu toate calităţile şi deficienţele acestuia.

Dacă în formula (25.47) se menţine constant nivelul caracteristicii se va obţine un indice unifactorial care să arate numai influenţa variaţiei caracteristicii şi poartă denumirea de *indice al variaţiei structurii*:

$$I_{t/o}^{f} = \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{0i}}{\sum_{i=1}^{n} f_{0i}} \qquad (25.49)$$

Între indicele bifactorial (25.47) şi indicii unifactoriali (25.48) şi (25.49), există o relaţie de sistem, bine definită şi anume: *produsul dintre cei doi indici unifactoriali este egal cu indicele bifactorial.*

Deci

$$I_{t/0}^{xf} = I_{t/0}^{x} \cdot I_{t/0}^{f} \qquad (25.50)$$

Substituind în (25.50) cei doi indici unifactoriali cu expresiile lor, se obţine:

$$\frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} x_{0i} f_{ti}} \cdot \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{0i}}{\sum_{i=1}^{n} f_{0i}}$$

Din relaţia (25.50) rezultă alte două relaţii utile în activitatea statistică:

1. *Indicele cu structură fixă se poate calcula şi ca un raport între indicele bifactorial şi indicele variaţiei structurii*:

$$I_{t/0}^{x} = I_{t/0}^{xf} : I_{t/0}^{f}$$

2. *Indicele variaţiei structurii este egal cu raportul dintre indicele bifactorial şi indicele cu structură fixă*:

$$I_{t/0}^{f} = I_{t/0}^{xf} \cdot I_{t/0}^{x}$$

Scanned with OKEN Scanner

Dacă în sistemul indicilor calculaţi ca raport între două medii aritmetice ponderate se substituie factorul $\frac{f_{ti}}{\sum_{i=1}^{n} f_{ti}} = G_{ti}$, atunci se obţine *forma agregat* a celor trei indici, astfel:

- *indicele bifactorial*:

$$I_{t/0}^{xf} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{0i}}{\sum_{i=1}^{n} f_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} G_{tj}}{\sum_{i=1}^{n} x_{0i} G_{0i}} \qquad (25.51)$$

- *indicele cu structură fixă*:

$$I_{t/0}^{x} = \frac{\sum_{i=1}^{n} x_{ti} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{ti}} = \frac{\sum_{i=1}^{n} x_{ti} G_{ti}}{\sum_{i=1}^{n} x_{0i} G_{ti}} \qquad (25.52)$$

- *indicele variaţiei structurii*:

$$I_{t/0}^{f} = \frac{\sum_{i=1}^{n} x_{0i} f_{ti}}{\sum_{i=1}^{n} f_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} f_{0i}}{\sum_{i=1}^{n} f_{0i}} = \frac{\sum_{i=1}^{n} x_{0i} G_{ti}}{\sum_{i=1}^{n} x_{0i} G_{0i}} \qquad (25.53)$$

Relaţia (25.50) se verifică şi în acest caz, astfel:

$$\frac{\sum_{i=1}^{n} x_{ti} G_{ti}}{\sum_{i=1}^{n} x_{0i} G_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} G_{ti}}{\sum_{i=1}^{n} x_{0i} G_{ti}} \cdot \frac{\sum_{i=1}^{n} x_{0i} G_{ti}}{\sum_{i=1}^{n} x_{0i} G_{0i}} \qquad (25.54)$$

Scanned with OKEN Scanner

Indicii calculaţi prin raportarea a două medii aritmetice ponderate se pot utiliza pentru studierea dinamicii oricăror fenomene. Mai des folosiţi sunt indicii productivităţii muncii, ai salariului mediu, ai costurilor de producţie, ai consumurilor lor specifice, ai nivelului relativ al cheltuielilor de circulaţie, ai recoltei medii la hectar, ai fenomenelor demografice (fertilitate, nupţialitate, mortalitate) etc.

Să considerăm, pentru exemplificare, situaţia producţiei dintr-un grup format din trei societăţi comerciale (tabelul 25.8).

Variaţia productivităţii zilnice a muncii, pe fiecare societate comercială în perioada curentă faţă de perioada de bază, este dată de indicii individuali ai productivităţii muncii (col. 7 a tabelului 25.8). Se observă că a avut loc o creştere a productivităţii muncii pe fiecare societate comercială din cadrul grupului şi anume, o creştere cu 33,3% în societatea *C*, cu 28% în societatea *A* şi cu 24,8% în societatea *B*.

La nivel de grup, dinamica productivităţii zilnice a muncii este dată de indicii de grup ai productivităţii muncii.

Astfel, variaţia productivităţii medii zilnice a muncii, datorată influenţei tuturor factorilor care au acţionat asupra ei, se determină cu ajutorul indicelui bifactorial al productivităţii muncii (25.47).

Tabelul 25.8

Societatea comercială	Perioada de bază		Perioada curentă		W_{0i}	W_{ti}	$i^{W}_{t/0}$	$W_{0i}T_{ti}$
	Volumul valoric al producţiei (mii lei)	Timpul de muncă cheltuit (mii om-zile)	Volumul valoric al producţiei (mii lei)	Timpul de muncă cheltuit (mii om-zile)				
0	1	2	3	4	5	6	7	8
A	5000	20	8000	25	250	320	1,28	6250
B	10000	40	9984	32	250	312	1,248	8000
C	4950	22	3000	10	225	300	1,333	2250
Total	19950	82	20984	67	243,3	313,2	1,287	16500

Substituind datele din tabelul 25.8 în formula (25.47), rezultă:

$$I^{W,T}_{t/0} = \frac{\sum_{i=1}^{n} W_{ti}T_{ti}}{\sum_{i=1}^{n} T_{ti}} : \frac{\sum_{i=1}^{n} W_{0i}T_{0i}}{\sum_{i=1}^{n} T_{0i}} = \frac{20984}{67} : \frac{19950}{82} = 1{,}287$$

$I_{t/0}^{W.T} = 1{,}287$ înseamnă că productivitatea zilnică a muncii pe total a crescut cu 28,7% în perioada curentă faţă de perioada de bază, ca urmare a creşterii productivităţii muncii pe fiecare societate şi a variaţiei structurii grupului de societăţi comerciale, din punct de vedere al timpului de muncă cheltuit pentru realizarea producţiei.

Influenţa separată a factorilor interni ai producţiei (mai buna organizare a procesului de producţie în societăţi comerciale, creşterea calificării muncitorilor, introducerea tehnicii noi, etc.) este dată de indicele unifactorial cu structură fixă al productivităţii muncii (25.48) care, aplicat la exemplul nostru, va fi:

$$I_{t/0}^{W} = \frac{\sum_{i=1}^{n} W_{ti} T_{ti}}{\sum_{i=1}^{n} W_{0i} T_{ti}} = \frac{20984}{16500} = 1{,}272$$

Deci, cea mai mare parte a creşterii productivităţii muncii la nivel de grup, şi anume 27,2%, s-a obţinut pe seama factorilor interni ai producţiei.

Cea de a doua categorie de factori care au acţionat asupra dinamicii productivităţii muncii pe total (factorii structurali sau *combinarea socială a procesului de producţie*) este dată de indicele unifactorial cu variaţie a structurii (25.49). Substituind datele din tabelul 25.8 în formula (25.49), rezultă:

$$I_{t/0}^{T} = \frac{\sum_{i=1}^{n} W_{0i} T_{ti}}{\sum_{i=1}^{n} T_{ti}} : \frac{\sum_{i=1}^{n} W_{0i} T_{0i}}{\sum_{i=1}^{n} T_{0i}} = \frac{16500}{67} : \frac{19950}{82} = 1{,}012$$

Aceasta înseamnă că factorii structurali au avut o influenţă mai mică asupra dinamicii productivităţii muncii pe total societate comercială şi anume, ei au determinat o creştere a acesteia de 1,2%.

Relaţia de sistem (25.50) se verifică astfel:

$$1{,}287 \cong 1{,}272 \cdot 1{,}012$$

Desigur, acelaşi rezultat s-ar fi obţinut şi dacă s-ar fi utilizat indicii sub formă agregat pentru calculul dinamicii productivităţii muncii pe total societăţi comerciale.

Scanned with OKEN Scanner

Calculele intermediare se sistematizează cu ajutorul tabelului 25.9.

Tabelul 25.9

Societatea comercială	T_{0i}	T_{ti}	W_{0i}	W_{ti}	$G_{0i}=\frac{T_{0i}}{\sum_{i=1}^{n}T_{0i}}$	$G_{ti}=\frac{T_{ti}}{\sum_{i=1}^{n}T_{ti}}$	$W_{0i}G_{0i}$	$W_{0i}G_{ti}$	$W_{ti}G_{ti}$
A	20	25	250	320	0,2439	0,3731	60,975	93,275	119,392
B	40	32	250	312	0,4878	0,4776	121,950	119,400	149,011
C	22	10	225	300	0,2683	0,1493	60,367	33,593	44,790
Total	82	67	243,3	313,2	1,0000	1,0000	243,292	246,268	313,193

Particularizând relaţia (25.51) pentru dinamica productivităţii muncii, rezultă formula indicelui agregat bifactorial al productivităţii muncii:

$$I_{t/0}^{W,G}=\frac{\sum_{i=1}^{n}W_{ti}G_{ti}}{\sum_{i=1}^{n}W_{0i}G_{0i}}=\frac{313,193}{243,292}=1,287$$

Indicele agregat cu structură fixă al productivităţii muncii va fi:

$$I_{t/0}^{W.G}=\frac{\sum_{i=1}^{n}W_{ti}G_{ti}}{\sum_{i=1}^{n}W_{0i}G_{ti}}=\frac{313,193}{246,268}=1,272$$

Indicele agregat al productivităţii muncii cu variaţie a structurii este dat de formula:

$$I_{t/0}^{G.W}=\frac{\sum_{i=1}^{n}W_{0i}G_{ti}}{\sum_{i=1}^{n}W_{0i}G_{0i}}=\frac{246,268}{243,292}=1,012$$

Scanned with OKEN Scanner

Observăm că s-au obţinut aceleaşi rezultate ca şi în cazul utilizării indicilor productivităţii muncii calculaţi din mărimi medii.

Utilizarea formei agregat a indicilor productivităţii muncii (şi, în general, a oricărui indice calculat din mărimi medii) dă posibilitatea determinării contribuţiei fiecărei societăţi comerciale la formarea nivelului mediu al productivităţii muncii, pe total societăţi comerciale. Astfel, în perioada curentă, productivitatea medie zilnică a muncii a fost de 313,193 mii lei pe un muncitor; la care societatea comercială *A* a contribuit cu 119,392 mii lei, societatea comercială *B*, cu 149,011 mii lei şi societatea comercială *C*, cu 44,79 mii lei.

Creşterea productivităţii medii zilnice a muncii pe total societate comercială, în perioada curentă faţă de perioada de bază, este dată de diferenţa dintre numărătorul şi numitorul indicelui agregat bifactorial, astfel:

$$\Delta_{t/0}^{W.G} = \sum_{i=1}^{n} W_{ti} G_{ti} - \sum_{i=1}^{n} W_{0i} G_{0i} = 313{,}193 - 243{,}292 = 69{,}90 \; lei/om - zi$$

Aceasta înseamnă că productivitatea medie zilnică pe un muncitor la nivel de grup de societăţi comerciale a crescut în perioada curentă faţă de perioada de bază cu 69,90 mii lei, ca urmare a influenţei ambelor categorii de factori.

Influenţa separată a celor două categorii de factori va fi:

- influenţa creşterii productivităţii muncii pe fiecare societate comercială a grupului în perioada curentă faţă de perioada de bază, dată de diferenţa dintre numărătorul şi numitorul formulei indicelui cu structură fixă, este:

$$\Delta_{t/0}^{W.G} = \sum_{i=1}^{n} W_{ti} G_{ti} - \sum_{i=1}^{n} W_{0i} G_{ti} = 313{,}193 - 246{,}268 = 66{,}93 \; lei/om - zi$$

ceea ce arată că factorii interni ai procesului de producţie din societăţi comerciale au avut o influenţă favorabilă asupra creşterii productivităţii zilnice a muncii la nivel de grup.

- influenţa factorilor structurali va fi dată de diferenţa dintre numărătorul şi numitorul formulei indicelui cu variaţii ale structurii, astfel:

$$\Delta_{t/0}^{G.W} = \sum_{i=1}^{n} W_{0i} G_{ti} - \sum_{i=1}^{n} W_{0i} G_{0i} = 246{,}268 - 243{,}292 = 2{,}98 \; lei$$

Desigur că, între influenţele factorilor productivităţii muncii, exprimate în mărimi absolute, se va verifica relaţia de sistem:

Scanned with OKEN Scanner

$$\Delta_{t/0}^{W,G} = \Delta_{t/0}^{W} + \Delta_{t/0}^{G}$$

25.4. Analiza influenţei factorilor cu ajutorul indicilor

25.4.1. Cu ajutorul indicilor agregaţi

Să considerăm un fenomen complex A care este rezultatul acţiunii a doi factori x şi y, legaţi între ei printr-o relaţie de produs :

$$A = x \cdot y$$

Influenţele celor doi factori se determină în mod obişnuit cu ajutorul indicilor agregaţi de tip *Laspeyres*, pentru un factor şi de tip *Paasche*, pentru celălalt.

Astfel:

- influenţa totală:

$$I_{t/0}^{A(x\cup y)} = \frac{\sum_{i=1}^{n} x_{ti}\, y_{ti}}{\sum_{i=1}^{n} x_{0i}\, y_{0i}}$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \sum_{i=1}^{n} x_{ti}\, y_{ti} - \sum_{i=1}^{n} x_{0i}\, y_{0i} \qquad (25.55)$$

- influenţa factorului x:

$$I_{t/0}^{A(x.y)} = \frac{\sum_{i=1}^{n} x_{ti}\, y_{0i}}{\sum_{i=1}^{n} x_{0i}\, y_{0i}}$$

şi

$$\Delta_{t/0}^{A(x.y)} = \sum_{i=1}^{n} x_{ti}\, y_{0i} - \sum_{i=1}^{n} x_{0i}\, y_{0i} \qquad (25.56)$$

Scanned with OKEN Scanner

• influenţa factorului *y*:

$$I_{t/0}^{A(y.x)} = \frac{\sum_{i=1}^{n} x_{ti}\, y_{ti}}{\sum_{i=1}^{n} x_{ti}\, y_{0i}}$$

şi

$$\Delta_{t/0}^{A(y.x)} = \sum_{i=1}^{n} x_{ti}\, y_{ti} - \sum_{i=1}^{n} x_{ti}\, y_{0i} \qquad (25.57)$$

Între influenţa totală şi cele două influenţe factoriale se verifică relaţiile:

1. *Produsul indicilor factoriali simpli este egal cu indicele factorial complex:*

$$I_{t/0}^{A(x.y)} \cdot I_{t/0}^{A(y.x)} = I_{t/0}^{A(x \cup y)} \qquad (25.58)$$

2. *Suma influenţelor factoriale simple exprimate în mărimi absolute este egală cu influenţa totală exprimată în aceleaşi unităţi de măsură:*

$$\Delta_{t/0}^{A(x \cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \qquad (25.59)$$

Aceste relaţii se verifică numai în cazul în care o influenţă factorială este calculată pe baza unui indice agregat de tip *Laspeyres* şi cealaltă, ca un indice de tip *Paasche*.

La rândul lor, influenţele totale ale fiecărui factor sunt formate din două componente:

• influenţa variaţiei volumului factorului respectiv;

• influenţa variaţiei structurii acestuia.

Pentru a separa aceste două componente, să considerăm pe rând cele două influenţe:

În cazul *influenţei totale a factorului x*, dată de relaţia (25.56), numărătorul acesteia se poate descompune în două componente astfel:

$$\sum_{i=1}^{n} x_{ti}\, y_{0i} = \sum_{i=1}^{n} x_{0i}\, y_{0i} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}} \qquad (25.60)$$

Scanned with OKEN Scanner

Substituind relaţia (25.60) în (25.56), rezultă indicele influenţei pure x_p a volumului factorului x:

$$I_{t/0}^{A(x_p \cdot y)} = \frac{\sum_{i=1}^{n} x_{0i} y_{0i} \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \qquad (25.61)$$

şi

$$\Delta_{t/0}^{A(x_p \cdot y)} = \sum_{i=1}^{n} x_{0i} y_{0i} \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}} - \sum_{i=1}^{n} x_{0i} y_{0i} = \sum_{i=1}^{n} x_{0i} y_{0i} \left(\frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}} - 1 \right)$$

- Componenta structurală x_s a factorului x va fi:

$$I_{t/0}^{A(x_s \cdot y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}} \qquad (25.62)$$

şi

$$\Delta_{t/0}^{A(x_s \cdot y)} = \sum_{i=1}^{n} x_{ti} y_{0i} - \sum_{i=1}^{n} x_{0i} y_{0i} \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}$$

Deci, influenţa totală a factorului x va fi:

- în mărimi relative:

$$I_{t/0}^{A(x \cdot y)} = I_{t/0}^{A(x_p \cdot y)} \cdot I_{t/0}^{A(x_s \cdot y)} \qquad (25.63)$$

Scanned with OKEN Scanner

- în mărimi absolute:

$$\Delta_{t/0}^{A(x.y)} = \Delta_{t/0}^{A(x_p.y)} + \Delta_{t/0}^{A(x_s.y)} \tag{25.64}$$

Analog se descompune influenţa totală a factorului y în cele două componente.

Influenţa pură y_p, a variaţiei factorului y asupra fenomenului complex A, este dată de indicele:

$$I_{t/0}^{A(y_p.x)} = \frac{\sum_{i=1}^{n} x_{ti}\, y_{0i} \dfrac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} y_{0i}}}{\sum_{i=1}^{n} x_{ti}\, y_{0i}} \tag{25.65}$$

şi

$$\Delta_{t/0}^{A(y_p.x)} = \sum_{i=1}^{n} x_{ti}\, y_{0i} \cdot \frac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} y_{0i}} - \sum_{i=1}^{n} x_{ti}\, y_{0i} = \sum_{i=1}^{n} x_{ti}\, y_{0i} \left(\frac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} y_{0i}} - 1 \right)$$

Influenţa structurală y_s a factorului y va fi dată de relaţia:

$$I_{t/0}^{A(y_s.x)} = \frac{\sum_{i=1}^{n} x_{ti}\, y_{ti}}{\sum_{i=1}^{n} x_{ti}\, y_{0i} \dfrac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} y_{0i}}} \tag{25.66}$$

şi

$$\Delta_{t/0}^{A(y_s.x)} = \sum_{i=1}^{n} x_{ti}\, y_{ti} - \sum_{i=1}^{n} x_{ti}\, y_{0i} \frac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} y_{0i}}$$

Rezultă că influenţa totală a factorului y va fi:

$$I_{t/0}^{A(y.x)} = I_{t/0}^{A(y_p.x)} \cdot I_{t/0}^{A(y_s.x)} \quad (25.67)$$

şi

$$\Delta_{t/0}^{A(y.x)} = \Delta_{t/0}^{A(y_p.x)} + \Delta_{t/0}^{A(y_s.x)} \quad (25.68)$$

In concluzie, descompunerea indicelui agregat al fenomenului complex pe cele două categorii de influenţe ale factorilor se prezintă schematic ca în figurile 25.1. şi 25.2.

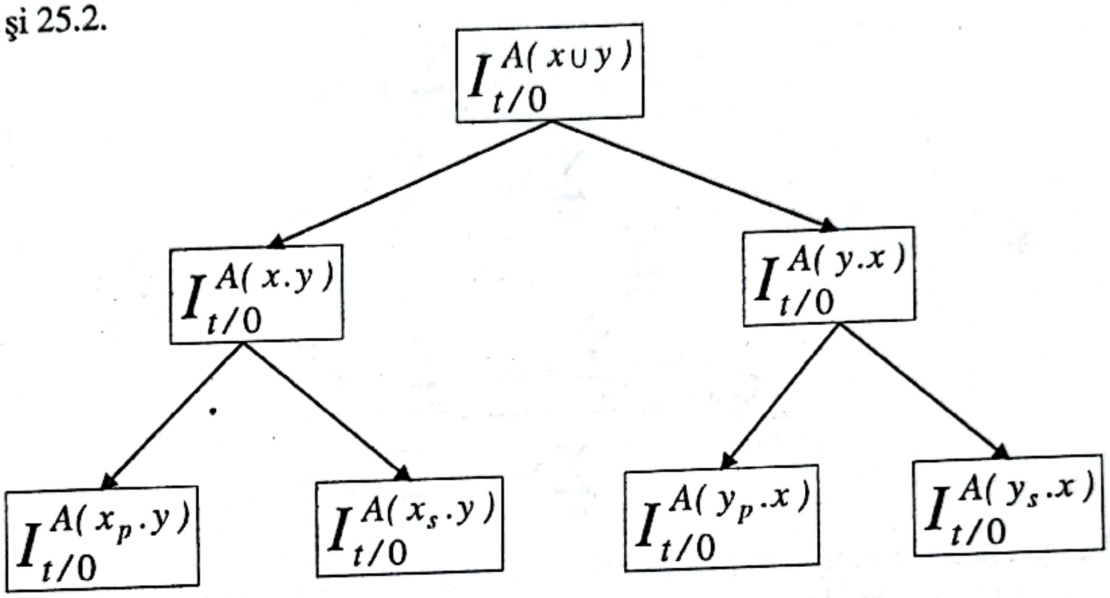


Figura 25.1. Schema analizei influenţelor relative ale factorilor în cazul indicilor agregaţi

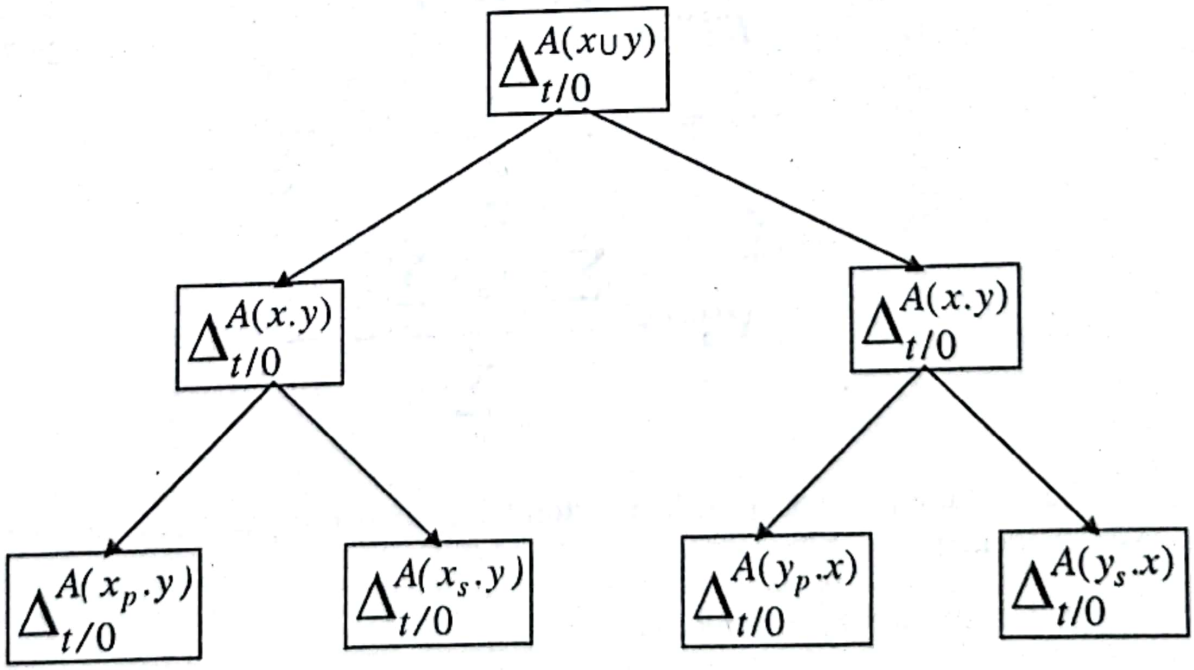


Figura 25.2. Schema analizei influenţelor absolute ale factorilor în cazul indicilor agregaţi

Scanned with OKEN Scanner

25.4.2. Cu ajutorul indicilor valorilor medii

Am văzut că dinamica valorilor medii a unor fenomene complexe se poate calcula cu ajutorul indicilor valorilor medii. Fie fenomenul complex $A = x \cdot y$, atunci indicele variaţiei totale a nivelului mediu pentru fenomenul x va fi:

$$I_{t/0}^{\bar{x}(x\cup y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} y_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} y_{0i}}{\sum_{i=1}^{n} y_{0i}} \qquad (25.69)$$

şi

$$\Delta_{t/0}^{\bar{x}(x\cup y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} y_{ti}} - \frac{\sum_{i=1}^{n} x_{0i} y_{0i}}{\sum_{i=1}^{n} y_{0i}}$$

care se poate descompune în doi factori:

a) variaţia totală a factorului x:

$$I_{t/0}^{\bar{x}(x.y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{ti}} \qquad (25.70)$$

$$\Delta_{t/0}^{\bar{x}(x,y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti} - \sum_{i=1}^{n} x_{0i} y_{ti}}{\sum_{i=1}^{n} y_{ti}}$$

La rândul ei, influenţa totală a factorului x (25.70) se poate descompune în două componente:

- variaţia pură x_p a factorului x:

Scanned with OKEN Scanner

$$I_{t/0}^{\bar{x}(x_p \cdot y)} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}}{\sum_{i=1}^{n} y_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} y_{ti}}{\sum_{i=1}^{n} y_{ti}} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}}{\sum_{i=1}^{n} x_{0i} y_{ti}} \qquad (25.71)$$

şi

$$\Delta_{t/0}^{\bar{x}(x_p \cdot y)} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti} \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}} - \sum_{i=1}^{n} x_{0i} y_{ti}}{\sum_{i=1}^{n} y_{ti}} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti} \left(\frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}} - 1 \right)}{\sum_{i=1}^{n} y_{ti}} \qquad (25.72)$$

• variaţia structurală x_s a factorului x va fi:

$$I_{t/0}^{\bar{x}(x_s \cdot y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} y_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} y_{ti} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}}{\sum_{i=1}^{n} y_{ti}} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{ti} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}} \qquad (25.73)$$

şi

Scanned with OKEN Scanner

$$\Delta_{t/0}^{\bar{x}(x_s \cdot y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti} - \sum_{i=1}^{n} x_{0i} y_{ti} \cdot \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i}}}{\sum_{i=1}^{n} y_{ti}}$$

Deci, între influenţa totală a factorului x şi cele două componente ale sale se verifică relaţiile:

$$I_{t/0}^{\bar{x}(x \cdot y)} = I_{t/0}^{\bar{x}(x_p \cdot y)} \cdot I_{t/0}^{\bar{x}(x_s \cdot y)} \tag{25.74}$$

şi

$$\Delta_{t/0}^{\bar{x}(x \cdot y)} = \Delta_{t/0}^{\bar{x}(x_p \cdot y)} + \Delta_{t/0}^{A(x_s \cdot y)}$$

b) variaţia factorului structural y va fi:

$$I_{t/0}^{\bar{x}(y \cdot x)} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti}}{\sum_{i=1}^{n} y_{ti}} : \frac{\sum_{i=1}^{n} x_{0i} y_{0i}}{\sum_{i=1}^{n} y_{0i}} \tag{25.75}$$

şi

$$\Delta_{t/0}^{\bar{x}(y \cdot x)} = \frac{\sum_{i=1}^{n} x_{0i} y_{ti}}{\sum_{i=1}^{n} y_{ti}} - \frac{\sum_{i=1}^{n} x_{0i} y_{0i}}{\sum_{i=1}^{n} y_{0i}}$$

Sintetizând, rezultă schemele generale de analiză a influenţei factorilor în cazul indicilor valorilor medii (fig 25.3. şi 25.4).

Scanned with OKEN Scanner

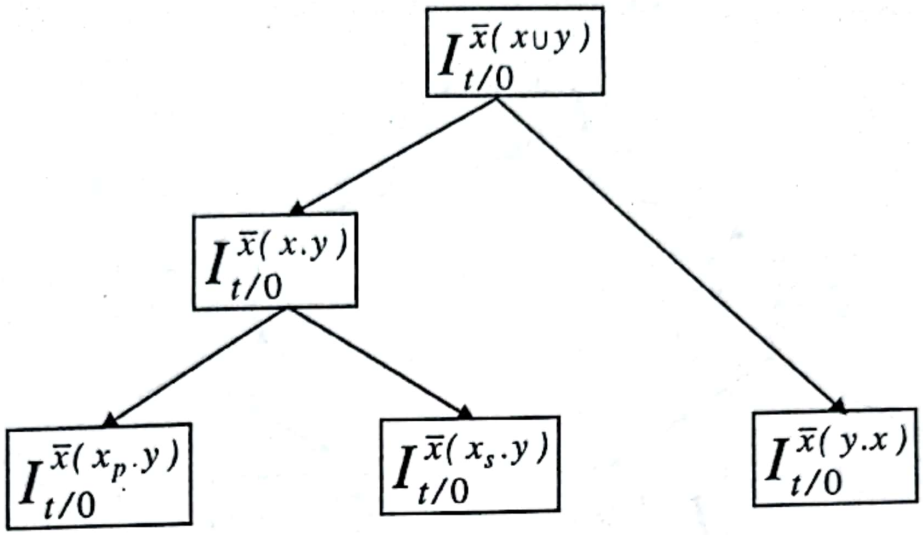


Figura 25.3. Schema analizei influenţelor relative ale factorilor în cazul indicilor valorilor medii

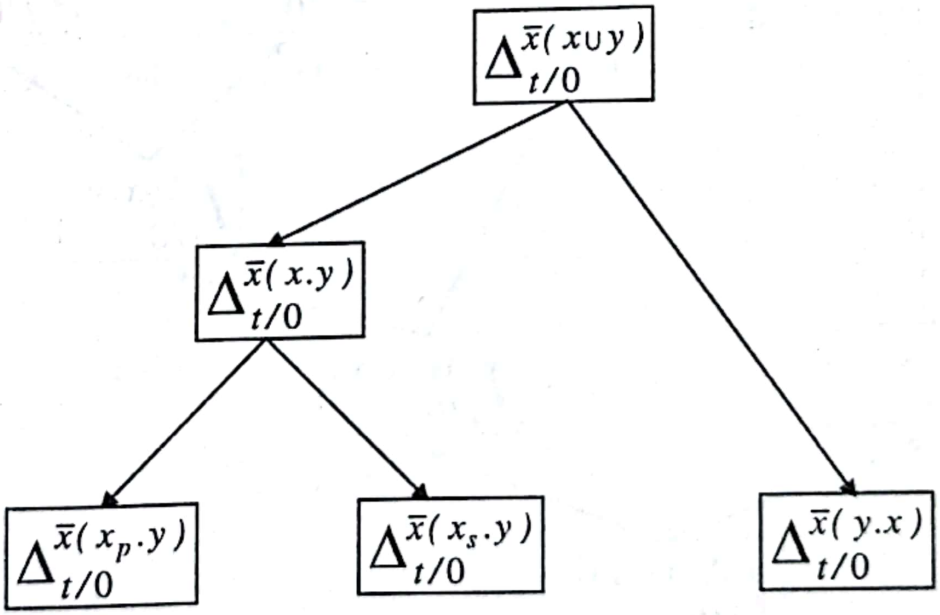
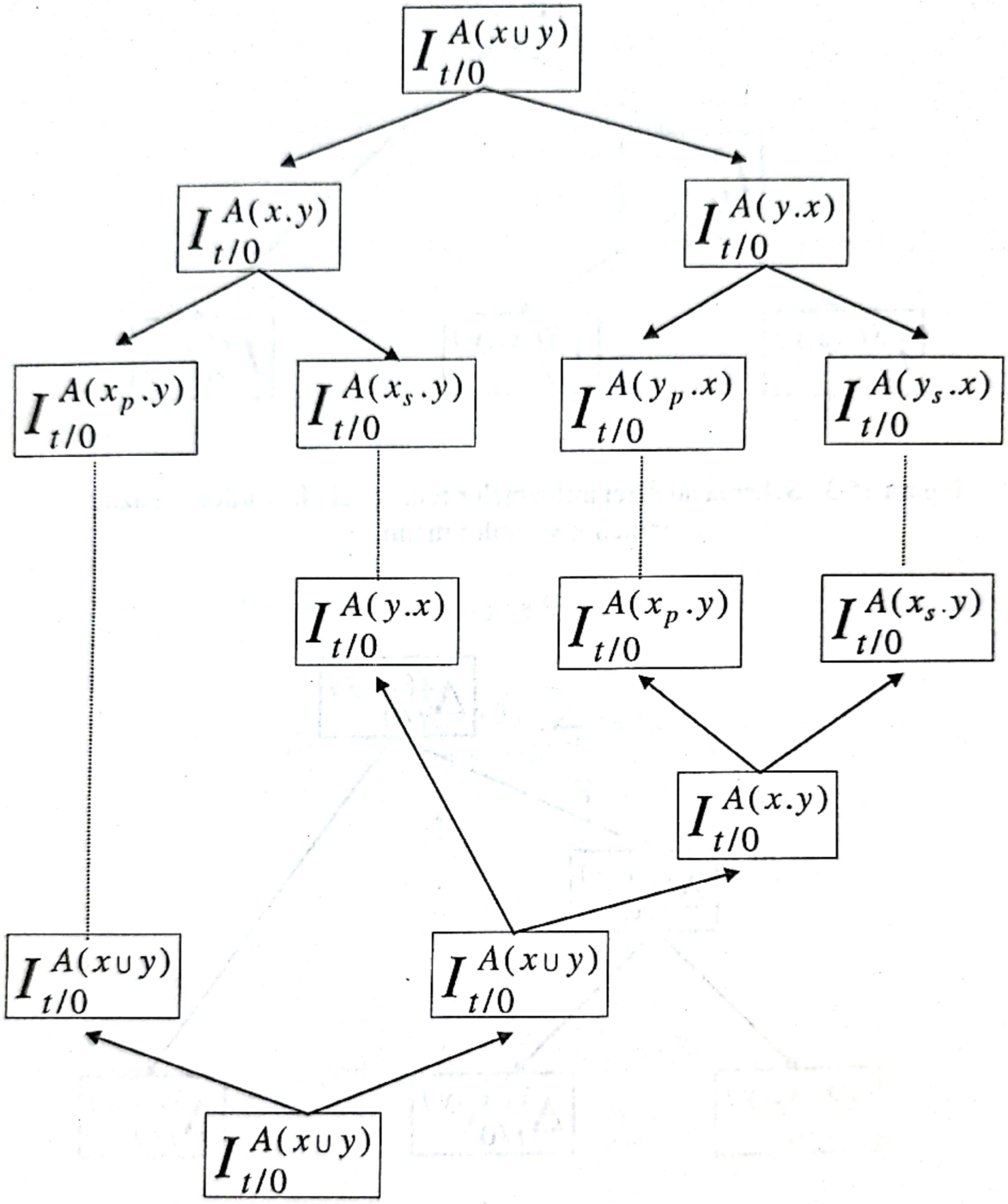


Figura 25.4. Schema analizei influenţelor absolute ale factorilor în cazul indicilor valorilor medii

Scanned with OKEN Scanner

Sistemul indicilor agregaţi



Sistemul indicilor valorilor medii

Fig. 25.5. Legătura dintre sistemul indicilor agregaţi şi sistemul indicilor valorilor medii

Scanned with OKEN Scanner

Între sistemul indicilor agregaţi şi sistemul indicilor valorilor medii există o strânsă legătură, care poate fi ilustrată de schema din fig.25.5:

25.5. Scheme de descompunere a indicilor pe factori

Descompunerea şi compunerea influenţei factorilor care acţionează asupra unui fenomen complex cu ajutorul indicilor, în funcţie de natura fenomenului şi de scopul analizei, se poate efectua cu ajutorul schemelor: *aditivă, multiplicativă şi mixtă (aditiv - multiplicativă).*

25.5.1. Schema aditivă

Fie un fenomen individual complex A care depinde de acţiunea a doi factori x şi y, legaţi între ei printr-o relaţie de însumare (aditivă):

$$A = x + y \tag{25.76}$$

Dinamica fenomenului complex A în perioada raportată t faţă de perioada de bază 0 va fi dată de indicele individual al fenomenului A:

- în mărimi aditive:

$$i_{t/0}^{A(x\cup y)} = \frac{A_t}{A_0} = \frac{x_t + y_t}{x_0 + y_0} \tag{25.77}$$

- în mărimi absolute:

$$\Delta_{t/0}^{A(x\cup y)} = A_t - A_0 = (x_t + y_t) - (x_0 + y_0) \tag{25.78}$$

Influenţele separate ale celor doi factori x şi y vor fi:

- în mărimi relative:

$$i_{t/0}^{A(x\cup y)} = \frac{x_t + y_t}{x_0 + y_0} = \frac{x_t}{x_0 + y_0} + \frac{y_t}{x_0 + y_0} = i_{t/0}^{A(x.y)} + i_{t/0}^{A(y.x)}$$

Scanned with OKEN Scanner

deci

$$i_{t/0}^{A(x\cup y)} = i_{t/0}^{A(x.y)} + i_{t/0}^{A(y.x)} \tag{25.79}$$

• în mărimi absolute:

$$\Delta_{t/0}^{A(x\cup y)} = (x_t + y_t) - (x_0 + y_0) = (x_t - x_0) + (y_t - y_0)$$

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{x} + \Delta_{t/0}^{y} \tag{25.80}$$

În cazul unui fenomen complex colectiv de forma: $\sum_{i=1}^{n} A_i = \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i$ cu ajutorul schemei aditive, se obţine:

• influenţa totală în mărimi relative:

$$I_{t/0}^{A} = \frac{\sum_{i=1}^{n} A_{ti}}{\sum_{i=1}^{n} A_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} + \sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} x_{0i} + \sum_{i=1}^{n} y_{0i}} =$$

$$= \frac{\sum_{i=1}^{n} x_{ti}}{\sum_{i=1}^{n} x_{0i} + \sum_{i=1}^{n} y_{0i}} + \frac{\sum_{i=1}^{n} y_{ti}}{\sum_{i=1}^{n} x_{0i} + \sum_{i=1}^{n} y_{0i}} = \tag{25.81}$$

$$= I_{t/0}^{A(x.y)} + I_{t/0}^{A(y.x)}$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \sum_{i=1}^{n} A_{ti} + \sum_{i=1}^{n} A_{0i} = \sum_{i=1}^{n} x_{ti} + \sum_{i=1}^{n} y_{ti} - \sum_{i=1}^{n} x_{0i} - \sum_{i=1}^{n} y_{0i} = \tag{25.82}$$

$$= (\sum_{i=1}^{n} x_{ti} - \sum_{i=1}^{n} x_{0i}) + (\sum_{i=1}^{n} y_{ti} - \sum_{i=1}^{n} y_{0i}) = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)}$$

Scanned with OKEN Scanner

Rezultă:

$$I_{t/0}^{A(x\cup y)} = I_{t/0}^{A(x.y)} + I_{t/0}^{A(y.x)} \tag{25.83}$$

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \tag{25.84}$$

Se observă că schema aditivă permite descompunerea în mod unic a variaţiei unui fenomen complex individual sau colectiv.

25.5.2. Schema multiplicativă

Schema multiplicativă se utilizează în analiza influenţei factorilor fenomenelor complexe individuale sau colective, care se prezintă sub formă de produs al factorilor:

$$A = \prod_{i=1}^{n} x_i \qquad i = \overline{1,n} \tag{25.85}$$

25.5.2.1. Metoda influenţelor separate *MIS*

În cazul unui fenomen individual complex, care depinde de doi factori x şi y, se verifică relaţia $A = x \cdot y$.

Dinamica fenomenului complex sub influenţa ambilor factori este dată de indicele individual al lui A:

- în mărimi relative:

$$i_{t/0}^{A(x\cup y)} = \frac{A_t}{A_0} = \frac{x_t y_t}{x_0 y_0} = \frac{x_t}{y_0} \cdot \frac{y_t}{y_0} = i_{t/0}^{x} \cdot i_{t/0}^{y} \tag{25.86}$$

deci

$$i_{t/0}^{A(x\cup y)} = i_{t/0}^{x} \cdot i_{t/0}^{y} \qquad i_{t/0}^{A(x\cup y)} = i_{t/0}^{x} \cdot i_{t/0}^{y} \tag{25.87}$$

- în mărimi absolute:

$$\Delta_{t/0}^{A(x\cup y)} = A_t - A_0 = x_t y_t - x_0 y_0 \tag{25.88}$$

Scanned with OKEN Scanner

Variaţia fenomenului individual complex A în perioada t faţă de perioada de bază se poate prezenta grafic cu ajutorul unui sistem de axe rectangulare (fig. 25.1).

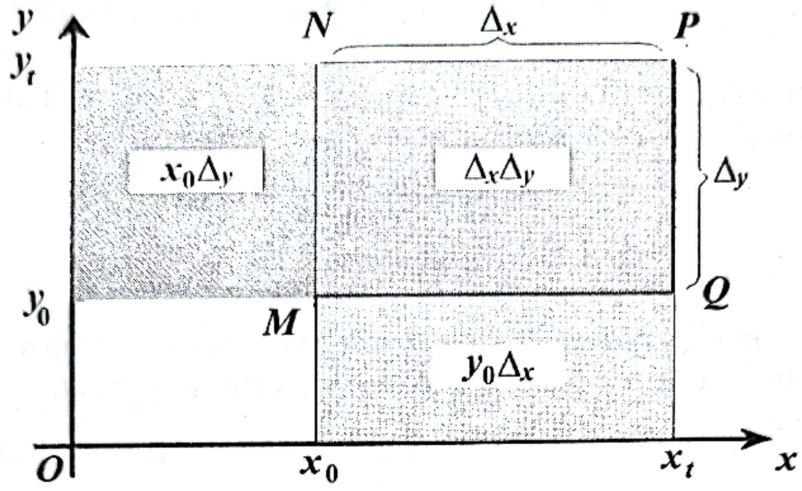


Figura 25.1. Variaţia în timp a unui fenomen individual complex bifactorial

Din figura 25.1 , rezultă:

$$\Delta_{t/0}^{A(x\cup y)} = x_t y_t - x_0 y_0 = Oy_t P x_t - Oy_0 M x_0 =$$
$$= y_0 y_t NM + x_0 MQ x_t + MNPQ =$$
$$= (y_t - y_0)x_0 + (x_t - x_0)y_0 + (y_t - y_0)(x_t - x_0)$$

deci

$$\Delta_{t/0}^{A(x\cup y)} = x_0\Delta_y + y_0\Delta_x + \Delta_x\Delta_y \qquad (25.89)$$

în care:

$\Delta_{t/0}^{A(x\cup y)}$ = variaţia totală în timp a fenomenului individual complex A pe seama ambilor factori x şi y ;

$x_0\Delta_y$ = variaţia în timp a fenomenului complex A determinată de influenţa separată a factorului y ;

$y_0\Delta_x$ = variaţia în timp a fenomenului complex A determinată de influenţa separată a factorului x ;

$\Delta_x\Delta_y$ = variaţia în timp a fenomenului complex A, determinată de influenţa concomitentă (interacţiunea) factorilor x şi y, numită şi *spor nedescompus.*

În acest caz, contribuţia factorilor la realizarea în timp a sporului total $\Delta_{t/0}^{A(x\cup y)}$ al fenomenului complex nu poate fi separată în mod unic. Din figura 25.6 rezultă că pot fi atribuite sigur suprafeţele y_0y_tNM şi x_0MQx_t factorului y şi respectiv x, fapt care nu este posibil pentru zona $MNPQ$.

Relaţia (25.89) arată că variaţia unui fenomen complex A, determinat de influenţa a doi factori x şi y, s-a descompus în trei influenţe factoriale.

Repartizarea influenţei comune pe cei doi factori x şi y se poate face convenţional în trei feluri:

1) se repartizează influenţa comună (interacţiunea) celor doi factori unui singur factor, de regulă celui calitativ;

2) se repartizează influenţa comună în mod egal pe cei doi factori x şi y:

- influenţa totală a factorului x va fi:

$$\Delta_{t/0}^{A(X.y)} = x_0\Delta_y + \frac{1}{2}\Delta_x\Delta_y \tag{25.90}$$

- influenţa totală a factorului y va fi:

$$\Delta_{t/0}^{A(Y.x)} = y_0\Delta_x + \frac{1}{2}\Delta_x\Delta_y \tag{25.91}$$

3) se repartizează influenţa comună pe factori, proporţional cu mărimea influenţelor lor separate:

- influenţa totală a factorului x:

$$\Delta_{t/0}^{A(X.y)} = y_0\Delta_x + \frac{y_0\Delta_x}{x_0\Delta_y + y_0\Delta_x}\Delta_x\Delta_y \tag{25.92}$$

- influenţa totală a factorului y:

$$\Delta_{t/0}^{A(Y.x)} = x_0\Delta_y + \frac{x_0\Delta_y}{x_0\Delta_y + y_0\Delta_x}\Delta_x\Delta_y \tag{25.93}$$

Scanned with OKEN Scanner

În cazul unor fenomene colective complexe de forma $\sum_{i=1}^{n} A_i = \sum_{i=1}^{n} x_i y_i$ prin aplicarea metodei influenţelor izolate MII, se obţine:

- influenţa totală a ambilor factori este dată de:

$$I_{t/0}^{A(x\cup y)} = \frac{\sum_{i=1}^{n} A_{ti}}{\sum_{i=1}^{n} A_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \tag{25.94}$$

Influenţele separate ale celor doi factori vor fi date de indicii unifactoriali corespunzători, condiţionaţi de menţinerea celuilalt factor la nivelul din perioada de bază:

- influenţa separată a factorului x:

$$I_{t/0}^{A(x/y_0)} = \frac{\sum_{i=1}^{n} x_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \tag{25.95}$$

- influenţa separată a factorului y:

$$I_{t/0}^{A(y/x_0)} = \frac{\sum_{i=1}^{n} y_{ti} x_{0i}}{\sum_{i=1}^{n} y_{0i} x_{0i}} \tag{25.96}$$

- influenţa simultană a celor doi factori se obţine eliminând din variaţia totală a ambilor factori produsul influenţelor separate ale factorilor:

$$I_{t/0}^{A(x\cap y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{0i}} : \left[\frac{\sum_{i=1}^{n} x_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \cdot \frac{\sum_{i=1}^{n} y_{ti} x_{0i}}{\sum_{i=1}^{n} y_{0i} x_{0i}} \right]. \tag{25.97}$$

sau

$$I_{t/0}^{A(x\cap y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{ti}} : \frac{\sum_{i=1}^{n} x_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \qquad (25.98)$$

Produsul acestor trei componente factoriale este egal cu indicele factorial complex:

$$I_{t/0}^{A(x\cup y)} = I_{t/0}^{A(x/y)} \cdot I_{t/0}^{A(y/x)} \cdot I_{t/0}^{A(x\cap y)} \qquad (25.99)$$

Cele trei componente factoriale exprimate în mărimi absolute se obţin ca diferenţă între numărătorul formulei (25.94) după substituirea $x_t = x_0 + \Delta_x$ şi $y_t = y_0 + \Delta_y$, astfel:

- influenţa totală a ambilor factori:

$$\begin{aligned}
\Delta_{t/0}^{A(x\cup y)} &= \sum_{i=1}^{n} x_{ti} y_{ti} - \sum_{i=1}^{n} x_{0i} y_{0i} = \\
&= \sum_{i=1}^{n} (x_{0i} + \Delta x_i)(y_{0i} + \Delta y_i) - \sum_{i=1}^{n} x_{0i} y_{0i} = \\
&= \sum_{i=1}^{n} x_{0i} y_{0i} + \sum_{i=1}^{n} x_{0i} \Delta y_i + \sum_{i=1}^{n} y_{0i} \Delta x_i + \sum_{i=1}^{n} \Delta x_i \Delta y_i - \sum_{i=1}^{n} x_{0i} y_{0i} = \\
&= \sum_{i=1}^{n} y_{0i} \Delta x_i + \sum_{i=1}^{n} x_{0i} \Delta y_i + \sum_{i=1}^{n} \Delta x_i \Delta y_i
\end{aligned} \qquad (25.100)$$

- influenţa separată a factorului x, pe baza relaţiei (25.95):

$$\Delta_{t/0}^{A(x/y_0)} = \sum_{i=1}^{n} x_{ti} y_{0i} - \sum_{i=1}^{n} x_{0i} y_{0i} = \sum_{i=1}^{n} y_{0i} \Delta x_i \qquad (25.101)$$

- influenţa separată a factorului y, pe baza relaţiei (25.96):

$$\Delta_{t/0}^{A(y/x_0)} = \sum_{i=1}^{n} y_{ti} x_{0i} - \sum_{i=1}^{n} y_{0i} x_{0i} = \sum_{i=1}^{n} x_{0i} \Delta y_i \qquad (25.102)$$

• influenţa modificării simultane a celor doi factori, pe baza relaţiei (25.98):

$$\Delta_{t/0}^{A(x\cap y)} = \sum_{i=1}^{n} x_{ti}y_{ti} - \sum_{i=1}^{n} x_{0i}y_{ti} - \sum_{i=1}^{n} x_{ti}y_{0i} + \sum_{i=1}^{n} x_{0i}y_{0i} =$$

$$= \sum_{i=1}^{n}(x_{ti} - x_{0i})y_{ti} - \sum_{i=1}^{n}(x_{ti} - x_{0i})y_{0i} = \qquad (25.103)$$

$$= \sum_{i=1}^{n}(x_{ti} - x_{0i})(y_{ti} - y_{0i}) = \sum_{i=1}^{n}\Delta x_i \Delta y_i$$

Suma celor trei componente factoriale este egală cu influenţa totală:

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{A(x/y_0)} + \Delta_{t/0}^{A(y/x_0)} + \Delta_{t/0}^{A(x\cap y)} \qquad (25.104)$$

Repartizarea influenţei comune pe factori se poate face pe baza aceloraşi ipoteze, ca şi în cazul fenomenelor individuale complexe.

a. Dacă influenţa comună se repartizează în întregime unui singur factor (de regulă celui calitativ), conduce la *M.S.L.*

b. În cazul repartizării influenţei comune în mod egal pe cei doi factori de influenţă, rezultă:

• influenţa totală a factorului x :

$$I_{t/0}^{A(x.y)} = \frac{\sum_{i=1}^{n} x_{ti}y_{0i}}{\sum_{i=1}^{n} x_{0i}y_{0i}} \cdot \sqrt{\frac{\sum_{i=1}^{n} x_{ti}y_{ti}}{\sum_{i=1}^{n} x_{0i}y_{ti}} : \frac{\sum_{i=1}^{n} x_{ti}y_{0i}}{\sum_{i=1}^{n} x_{0i}y_{0i}}}$$

sau

$$I_{t/0}^{A(x.y)} = I_{t/0}^{A(x/y_0)} \cdot I_{t/0}^{A(x\cap y)} \qquad (25.105)$$

şi

$$\Delta_{t/0}^{A(x.y)} = \sum_{i=1}^{n} y_{0i}\Delta_{x_i} + \frac{\sum_{i=1}^{n} \Delta_{x_i}\Delta_{y_i}}{2}$$

Scanned with OKEN Scanner

sau

$$\Delta_{t/0}^{A(x.y)} = \Delta_{t/0}^{A(x/y_0)} + \frac{\Delta_{t/0}^{A(x\cap y)}}{2} \tag{25.106}$$

• influenţa totală a factorului y :

$$I_{t/0}^{A(y.x)} = \frac{\sum_{i=1}^{n} y_{ti}x_{0i}}{\sum_{i=1}^{n} y_{0i}x_{0i}} \cdot \sqrt{\frac{\sum_{i=1}^{n} x_{ti}y_{ti}}{\sum_{i=1}^{n} x_{0i}y_{ti}} : \frac{\sum_{i=1}^{n} x_{ti}y_{0i}}{\sum_{i=1}^{n} x_{0i}y_{0i}}}$$

sau

$$I_{t/0}^{A(y.x)} = I_{t/0}^{A(y/x_0)} \cdot \sqrt{I_{t/0}^{A(x\cap y)}} \tag{25.107}$$

şi

$$\Delta_{t/0}^{A(y.x)} = \sum_{i=1}^{n} x_{0i}\Delta_{y_i} + \frac{\sum_{i=1}^{n} \Delta_{x_i}\Delta_{y_i}}{2}$$

sau

$$\Delta_{t/0}^{A(y.x)} = \Delta_{t/0}^{A(y/x_0)} + \frac{\Delta_{t/0}^{A(x\cap y)}}{2} \tag{25.108}$$

Suma (respectiv produsul) influenţelor factoriale totale este egală cu variaţia totală a fenomenului complex:

$$I_{t/0}^{A(x\cup y)} = I_{t/0}^{A(x.y)} \cdot I_{t/0}^{A(y.x)} \tag{25.109}$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \tag{25.110}$$

c. În ipoteza repartizării influenţei comune pe factori, proporţional cu influenţa lor separată, se obţine:

• influenţa totală a factorului x :

$$I_{t/0}^{A(x.y)} = I_{t/0}^{A(x/y_0)} \cdot \sqrt{\frac{I_{t/0}^{A(x/y_0)}}{I_{t/0}^{A(y/x_0)}} \cdot I_{t/0}^{A(x\cap y)}} \tag{25.111}$$

şi

$$\Delta_{t/0}^{A(x.y)} = \Delta_{t/0}^{A(x/y_0)} + \frac{\Delta_{t/0}^{A(x/y_0)}}{\Delta_{t/0}^{A(x/y_0)} + \Delta_{t/0}^{A(y/x_0)}} \cdot \Delta_{t/0}^{A(x\cap y)} \tag{25.112}$$

• influenţa totală a factorului y :

$$I_{t/0}^{A(y.x)} = I_{t/0}^{A(y/x_0)} \cdot \sqrt{\frac{I_{t/0}^{A(x/y_0)}}{I_{t/0}^{A(y/x_0)}} \cdot I_{t/0}^{A(x\cap y)}} \tag{25.113}$$

şi

$$\Delta_{t/0}^{A(y.x)} = \Delta_{t/0}^{A(y/x_0)} + \frac{\Delta_{t/0}^{A(y/x_0)}}{\Delta_{t/0}^{A(x/y_0)} + \Delta_{t/0}^{A(y/x_0)}} \cdot \Delta_{t/0}^{A(x\cap y)} \tag{25.114}$$

Între influenţele totale ale factorilor şi variaţia fenomenului complex, se verifică relaţiile:

$$I_{t/0}^{A(x\cup y)} = I_{t/0}^{A(x.y)} \cdot I_{t/0}^{A(y.x)} \tag{25.115}$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \tag{25.116}$$

În cazul unui fenomen individual complex determinat de trei factori $A = x \cdot y \cdot z$, dinamica lui A în perioada raportată t faţă de perioada de bază 0 va fi dată de indicele individual:

$$i_{t/0}^{A(x\cup y\cup z)} = \frac{A_t}{A_0} = \frac{x_t y_t z_t}{x_0 y_0 z_0} = i_{t/0}^{x} \cdot i_{t/0}^{y} \cdot i_{t/0}^{z} \tag{25.117}$$

şi

$$\Delta_{t/0}^{A(x\cup y\cup z)} = A_t - A_0 = x_t y_t z_t - x_0 y_0 z_0 \tag{25.118}$$

În intervalul de timp $(0 - t)$, considerăm:

$$\begin{aligned} x_t &= x_0 + \Delta_x \\ y_t &= y_0 + \Delta_y \\ x_t &= z_0 + \Delta_z \end{aligned} \tag{25.119}$$

Scanned with OKEN Scanner

Ţinând seama de relaţiile (25.119), relaţia (25.118) devine:

$$\Delta_{t/0}^{A(x\cup y\cup z)} = (x_0+\Delta_x)(y_0+\Delta_y)(z_0+\Delta_z) - x_0y_0z_0 =$$

$$= (x_0y_0 + x_0\Delta_y + y_0\Delta_x + \Delta_x\Delta_y)(z_0+\Delta_z) - x_0y_0z_0 =$$

$$= x_0y_0z_0 + x_0y_0\Delta_z + x_0z_0\Delta_y + x_0\Delta_y\Delta_z + y_0z_0\Delta_x +$$

$$+ y_0\Delta_x\Delta_z + z_0\Delta_x\Delta_y + \Delta_x\Delta_y\Delta_z - x_0y_0z_0$$

Deci, variaţia în timp a fenomenului individual complex s-a descompus în 7 influenţe, din care : *trei influenţe separate* corespunzătoare celor trei factori, *trei influenţe simultane a câte doi factori* şi *o influenţă simultană a tuturor factorilor*.

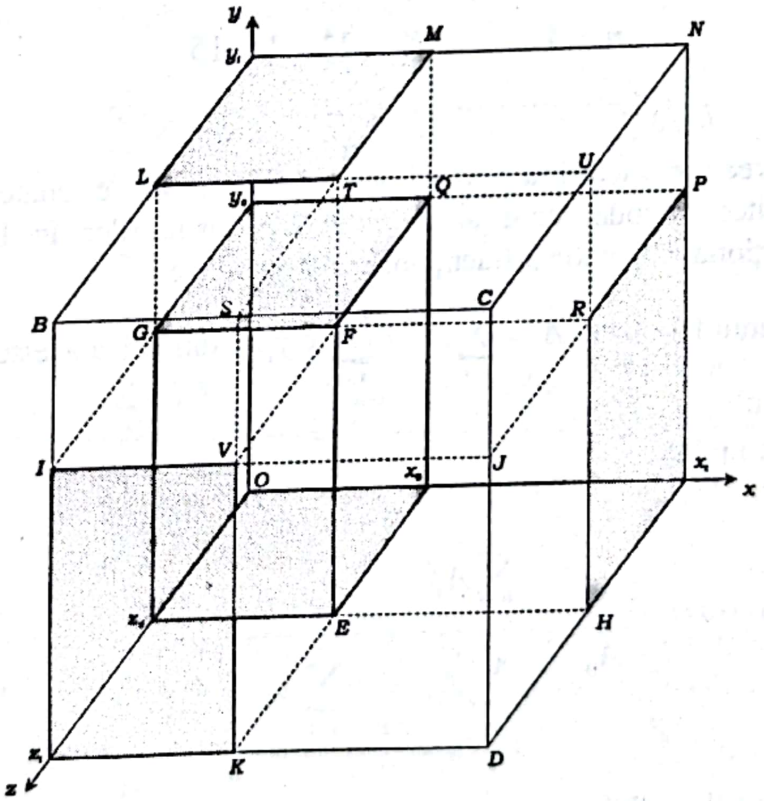


Figura 25.2. Variaţia în timp a unui fenomen individual complex trifactorial

Scanned with OKEN Scanner

La acelaşi rezultat al descompunerii pe factori a variaţiei fenomenului complex A, se mai poate ajunge şi pe cale grafică (fig. 25.2).

Cele şapte influenţe se prezintă sub forma unor paralelipipede dreptunghiulare.

$$\Delta_{t/0}^{A(xUyUz)} = x_t y_t z_t - x_0 y_0 z_0 = V(Ox_t Dz_t BCNy_t) - \\ - V(Ox_0 Ez_0 Gy_0 QF) = \Delta_x y_0 z_0 + \Delta_y x_0 z_0 + \Delta_z x_0 y_0 + \\ + \Delta_x \Delta_y z_0 + \Delta_x \Delta_z y_0 + \Delta_y \Delta_z x_0 + \Delta_x \Delta_y \Delta_z \qquad (25.120)$$

În general, numărul influenţelor factoriale K rezultate din descompunerea variaţiei totale în timp a fenomenelor complexe care depind de n factori este dat de relaţia:

$$K = 2^n - 1 \qquad (25.121)$$

pentru

$n = 2$ $\quad K = 2^2 - 1 = 3$

$n = 3$ $\quad K = 2^3 - 1 = 7$

$n = 4$ $\quad K = 2^4 - 1 = 15$

.....

Repartizarea pe factori a sporurilor nedescompuse se poate efectua cu ajutorul mai multor metode, cum ar fi: metoda substituirilor în lanţ, metoda repartizării proporţionale, metoda influenţelor izolate şi altele.

În cazul unui fenomen $A = \sum_{i=1}^{n} A_i = \sum_{i=1}^{n} x_i y_i z_i$ dinamica acestuia va fi dată de indicele agregat:

- în mărimi relative:

$$I_{t/0}^{A(x,y,z)} = \frac{A_t}{A_0} = \frac{\sum_{i=1}^{n} A_{ti}}{\sum_{i=1}^{n} A_{0i}} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti} z_{ti}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} \qquad (25.122)$$

- în mărimi absolute:

$$\Delta_{t/0}^{A(x,y,z)} = \sum_{i=1}^{n} x_{ti} y_{ti} z_{ti} - \sum_{i=1}^{n} x_{0i} y_{0i} z_{0i} \qquad (25.123)$$

sau

$$\Delta_{t/0}^{A(x\cup y\cup z)} = \sum_{i=1}^{n} (x_{0i} + \Delta_{x_i})(y_{0i} + \Delta_{y_i})(z_{0i} + \Delta_{z_i}) - \sum_{i=1}^{n} x_{0i} y_{0i} z_{0i} =$$

$$= \sum_{i=1}^{n} x_{0i} y_{0i} \Delta_{z_i} + \sum_{i=1}^{n} x_{0i} z_{0i} \Delta_{y_i} + \sum_{i=1}^{n} x_{0i} \Delta_{y_i} \Delta_{z_i} + \sum_{i=1}^{n} y_{0i} z_{0i} \Delta_{x_i} +$$

$$+ \sum_{i=1}^{n} y_{0i} \Delta_{x_i} \Delta_{z_i} + \sum_{i=1}^{n} z_{0i} \Delta_{x_i} \Delta_{y_i} + \sum_{i=1}^{n} \Delta_{x_i} \Delta_{y_i} \Delta_{z_i}$$

Variaţia relativă a fenomenului complex colectiv va fi dată de următorii indici agregaţi:

- influenţele separate ale factorilor:

$$I_{t/0}^{A(x,y,z)} = \frac{\sum_{i=1}^{n} x_{ti} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \qquad (25.124)$$

$$I_{t/0}^{A(x,y,z)} = \frac{\sum_{i=1}^{n} y_{ti} x_{0i} z_{0i}}{\sum_{i=1}^{n} y_{0i} x_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \qquad (25.125)$$

$$I_{t/0}^{A(x,y,z)} = \frac{\sum_{i=1}^{n} z_{ti} x_{0i} y_{0i}}{\sum_{i=1}^{n} z_{0i} x_{0i} y_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \qquad (25.126)$$

- influenţele simultane ale perechilor de factori:

Scanned with OKEN Scanner

$$I_{t/0}^{A(x\cap y,z)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} \tag{25.127}$$

sau

$$I_{t/0}^{A(x\cap y,z)} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} i_{t/0}^{y_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} =$$

$$= \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}}$$

deci

$$I_{t/0}^{A(x\cap y,z)} = I_{t/0}^{A(x,y,z)} \cdot I_{t/0}^{A(y,x,z)} \tag{25.128}$$

şi

$$I_{t/0}^{A(x\cap z,y)} = I_{t/0}^{A(x,y,z)} \cdot I_{t/0}^{A(z,x,y)} \tag{25.129}$$

sau

$$I_{t/0}^{A(x\cap z,y)} = \frac{\sum_{i=1}^{n} x_{ti} z_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} z_{0i} y_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} =$$

$$= \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}}$$

şi

$$I_{t/0}^{A(y\cap z,x)} = I_{t/0}^{A(y,x,z)} \cdot I_{t/0}^{A(z,x,y)} \tag{25.130}$$

sau

$$I_{t/0}^{A(y\cap z,y)} = \frac{\sum_{i=1}^{n} y_{ti} z_{ti} x_{0i}}{\sum_{i=1}^{n} y_{0i} z_{0i} x_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} =$$

$$= \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}}$$

Influenţa simultană a tuturor factorilor va fi:

$$I_{t/0}^{A(x\cap y\cap z)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti} z_{ti}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} i_{t/0}^{y_i} i_{t/0}^{z_i} x_{0i} y_{0i} z_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i} z_{0i}} = \tag{25.131}$$

$$= \frac{\sum_{i=1}^{n} i_{t/0}^{x_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{y_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}} \cdot \frac{\sum_{i=1}^{n} i_{t/0}^{z_i} A_{0i}}{\sum_{i=1}^{n} A_{0i}}$$

Analog probabilităţilor condiţionate, se pot defini şi indicii dinamicii unui fenomen complex colectiv, al unui grup de factori, condiţionat de un alt grup de factori, ca un raport dintre indicele tuturor factorilor şi indicele factorilor care condiţionează, astfel:

$$I_{t/0}^{A(x,z/y)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(y,x,z)}} \tag{25.132}$$

$$I_{t/0}^{A(x,y/z)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(z,x,y)}} \tag{25.133}$$

Scanned with OKEN Scanner

$$I_{t/0}^{A(y,z/x)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(x,y,z)}} \qquad (25.134)$$

şi

$$I_{t/0}^{A(x/y,z)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(y\cap z,x)}} \qquad (25.135)$$

$$I_{t/0}^{A(y/x,z)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(x\cap z,y)}} \qquad (25.136)$$

$$I_{t/0}^{A(z/x,y)} = \frac{I_{t/0}^{A(x,y,z)}}{I_{t/0}^{A(x\cap y,z)}} \qquad (25.137)$$

25.5.2.2. Metoda substituirii în lanţ

Metoda substituirilor în lanţ MSL porneşte de la ipoteza acceptării unei anumite succesiuni convenţionale în exercitarea şi separarea acţiunii factorilor şi, în acest caz , sporul nedescompus se atribuie în întregime unuia dintre factori. Pornind de la ipoteza că în dezvoltarea fenomenelor au loc în primul rând variaţii cantitative şi în al doilea rând cele calitative, se consideră în calcule, în cadrul fenomenelor complexe, această succesiune.

Principiul de bază al metodei *MSL* constă în considerarea, în cazul unui fenomen dependent de n factori, a unui singur factor variabil şi $n-1$ factori invariabili, apoi a doi factori variabili şi $n-2$ factori invariabili şi aşa mai departe. Prin utilizarea acestei metode, sporul total al fenomenului complex se separă în întregime pe factori, iar numărul indicilor factoriali ai fenomenului complex este egal cu numărul factorilor luaţi în calcul.

Fie un fenomen individual complex:

$$A = x_1 \cdot x_2 \cdot \cdots \cdot x_i \cdot \cdots \cdot x_n = \prod_{i=1}^{n} x_i \qquad i = \overline{1,n} \qquad (25.138)$$

a cărei dinamică este dată de indicele individual:

• în mărimi relative:

$$i_{t/0}^{A(x_i)} = \frac{A_t}{A_0} = \frac{x_{t1} \cdot x_{t2} \cdots x_{ti} \cdots x_{tn}}{x_{01} \cdot x_{02} \cdots x_{0i} \cdots x_{0n}} = \frac{\prod_{i=1}^{n} x_{ti}}{\prod_{i=1}^{n} x_{0i}} \qquad i = \overline{1,n}$$

sau

$$i_{t/0}^{A(x_i)} = \frac{x_{t1}}{x_{01}} \cdot \frac{x_{t2}}{x_{02}} \cdots \frac{x_{tn}}{x_{0n}} = i_{t/0}^{x_1} \cdot i_{t/0}^{x_2} \cdots i_{t/0}^{x_n}$$

deci

$$i_{t/0}^{A(x_i)} = \prod_{i=1}^{n} i_{t/0}^{x_i} \qquad i = \overline{1,n} \qquad (25.139)$$

• în mărimi absolute:

$$\Delta_{t/0}^{A(x_i)} = A_t - A_0 = x_{t1} \cdot x_{t2} \ldots x_{tn} - x_{01} \cdot x_{02} \ldots x_{0n} =$$

$$= (x_{t1} - x_{01})(x_{t2} - x_{02}) \ldots (x_{tn} - x_{0n})$$

sau

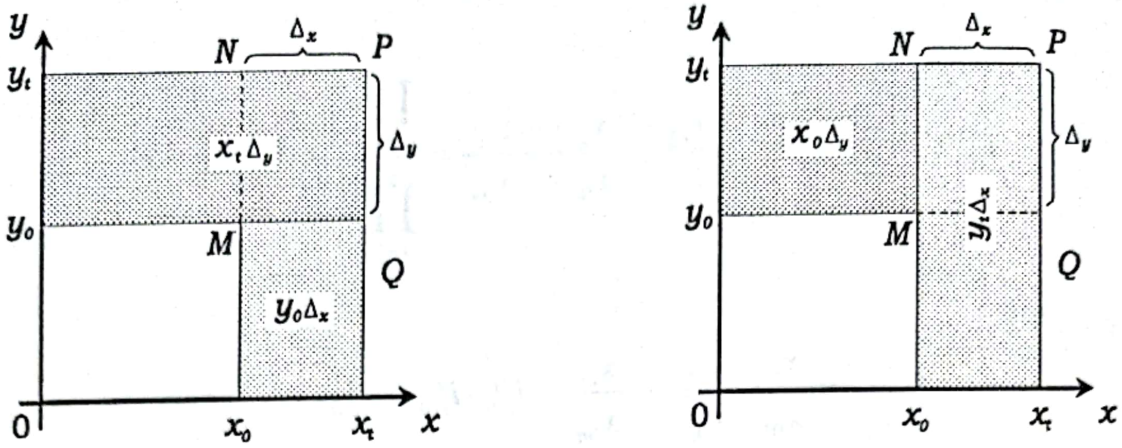
$$\Delta_{t/0}^{A(x_i)} = \Delta_{t/0}^{A(x_1)} \cdot \Delta_{t/0}^{A(x_2)} \cdots \Delta_{t/0}^{A(x_n)} \qquad (25.140)$$

În continuare, se va particulariza metoda substituirilor în lanţ pentru cazul fenomenelor complexe formate din doi şi trei factori.

Pentru $n = 2$ $A = x \cdot y$.

Variaţia în timp a fenomenului A se poate ilustra grafic cu ajutorul figurii 25.3.

Scanned with OKEN Scanner



Varianta I

Varianta a II a

Figura 25.3. Variante ale metodei substituirilor în lanţ
în funcţie de ordinea de substituire

Dinamica fenomenului individual complex va fi dată de indicele individual:

- în mărimi relative:

$$i_{t/0}^{A(x\cup y)} = \frac{A_t}{A_0} = \frac{x_t \cdot y_t}{x_0 \cdot y_0} = i_{t/0}^{x} \cdot i_{t/0}^{y} \qquad (25.141)$$

- în mărimi absolute:

$$\Delta_{t/0}^{A(x\cup y)} = A_t - A_0 = x_t \cdot y_t - x_0 \cdot y_0 \qquad (25.142)$$

Influenţa fiecărui factor asupra variaţiei fenomenului complex depinde de ordinea substituirii, deci vor exista două variante (fig. 25.9).

În *varianta a I^a*, se substituie întâi factorul x:

- influenţa factorului x va fi:

$$i_{t/0}^{A(x.y)} = \frac{x_t \cdot y_0}{x_0 \cdot y_0} = i_{t/0}^{x}$$

şi

$$\Delta_{t/0}^{A(x.y)} = (x_t - x_0)y_0 = y_o\Delta_x \qquad (25.143)$$

• influenţa factorului y va fi:

$$i_{t/0}^{A(y.x)} = \frac{x_t \cdot y_t}{x_t \cdot y_0} = i_{t/0}^{y}$$

şi

$$\Delta_{t/0}^{A(y.x)} = x_t (y_t - y_0) = x_t \Delta_y \qquad (25.144)$$

Între variaţia fenomenului complex şi variaţiile factoriale se verifică relaţiile:

$$i_{t/0}^{A(x \cup y)} = i_{t/0}^{x} \cdot i_{t/0}^{y} \qquad (25.145)$$

şi

$$\Delta_{t/0}^{A(x \cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \qquad (25.146)$$

În condiţiile variantei a I^a de substituire, întreg sporul nedescompus $MNPQ$, din figura 25.9, se atribuie factorului y.

În *varianta a* II^a, se substituie întâi factorul y şi deci, întreg sporul nedescompus se atribuie lui x:

• influenţa factorului y:

$$i_{t/0}^{A(y.x)} = \frac{y_t \cdot x_0}{y_0 \cdot x_0} = i_{t/0}^{y}$$

şi

$$\Delta_{t/0}^{A(y.x)} = (y_t - y_0) x_0 = x_0 \Delta_y \qquad (25.147)$$

• influenţa factorului x:

$$i_{t/0}^{A(x.y)} = \frac{y_t \cdot x_t}{y_t \cdot x_0} = i_{t/0}^{x}$$

şi

$$\Delta_{t/0}^{A(x.y)} = y_t (x_t - x_0) = y_t \Delta_x \qquad (25.148)$$

Şi în acest caz se verifică relaţiile de sistem:

$$i_{t/0}^{A(x \cup y)} = i_{t/0}^{A(y.x)} \cdot i_{t/0}^{A(x.y)} \qquad (25.149)$$

şi

$$\Delta_{t/0}^{A(x \cup y)} = \Delta_{t/0}^{A(y.x)} + \Delta_{t/0}^{A(x.y)} \qquad (25.150)$$

Scanned with OKEN Scanner

Din ambele variante, rezultă că, odată cu schimbarea ordinii de substituire a factorilor, se schimbă şi mărimea contribuţiei lor la variaţia fenomenului complex. Un factor substituit rămâne cu valoarea din perioada raportată atât la numărătorul, cât şi la numitorul indicelui, în toate substituţiile ulterioare ale factorilor.

În cazul fenomenelor colective complexe, formate sub influenţa a doi factori, aplicarea metodei substituirilor în lanţ pentru determinarea influenţei factorilor se poate efectua tot în două variante, în funcţie de ordinea de substituire.

Astfel, fie $A = \sum_{i=1}^{n} x_i y_i$ atunci:

- variaţia totală:

$$I_{t/0}^{A(x\cup y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{0i} y_{0i}}$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \sum_{i=1}^{n} x_{ti} y_{ti} - \sum_{i=1}^{n} x_{0i} y_{0i} \qquad (25.151)$$

- variaţia factorială:

varianta a I^a *varianta a II^a*

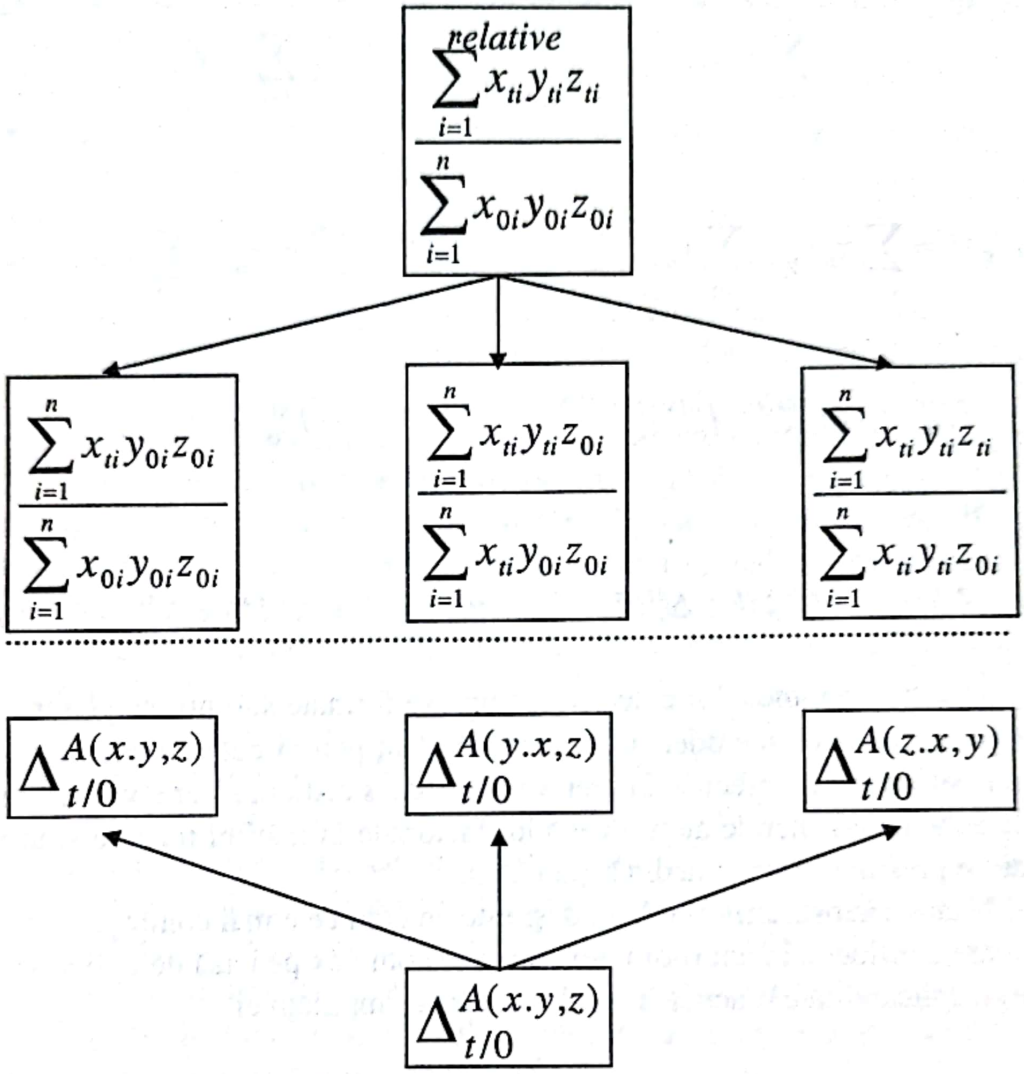
$$I_{t/0}^{A(x.y)} = \frac{\sum_{i=1}^{n} x_{ti} y_{0i}}{\sum_{i=1}^{n} x_{0i} y_{0i}} \qquad I_{t/0}^{A(y.x)} = \frac{\sum_{i=1}^{n} y_{ti} x_{0i}}{\sum_{i=1}^{n} y_{0i} x_{0i}} \qquad (25.152)$$

şi

$$\Delta_{t/0}^{A(x.y)} = \sum_{i=1}^{n} x_{ti} y_{0i} - \sum_{i=1}^{n} x_{0i} y_{0i} \qquad \Delta_{t/0}^{A(y.x)} = \sum_{i=1}^{n} y_{ti} x_{0i} - \sum_{i=1}^{n} y_{0i} x_{0i} \qquad (25.153)$$

iar

Mărimi relative



Mărimi absolute

Figura 25.9. Schema influenţelor factoriale în cazul fenomenelor $A = x \cdot y \cdot z$, după *MSL*

Scanned with OKEN Scanner

$$I_{t/0}^{A(y.x)} = \frac{\sum_{i=1}^{n} x_{ti} y_{ti}}{\sum_{i=1}^{n} x_{ti} y_{0i}} \qquad I_{t/0}^{A(x.y)} = \frac{\sum_{i=1}^{n} y_{ti} x_{ti}}{\sum_{i=1}^{n} y_{ti} x_{0i}} \tag{25.154}$$

şi

$$\Delta_{t/0}^{A(y.x)} = \sum_{i=1}^{n} x_{ti} y_{ti} - \sum_{i=1}^{n} x_{ti} y_{0i} \qquad \Delta_{t/0}^{A(x.y)} = \sum_{i=1}^{n} y_{ti} x_{ti} - \sum_{i=1}^{n} y_{ti} x_{0i} \tag{25.155}$$

$$I_{t/0}^{A(x\cup y)} = I_{t/0}^{A(x.y)} \cdot I_{t/0}^{A(y.x)} \qquad I_{t/0}^{A(y\cup x)} = I_{t/0}^{A(y.x)} \cdot I_{t/0}^{A(x.y)} \qquad 25.156)$$

şi

$$\Delta_{t/0}^{A(x\cup y)} = \Delta_{t/0}^{A(x.y)} + \Delta_{t/0}^{A(y.x)} \qquad \Delta_{t/0}^{A(y\cup x)} = \Delta_{t/0}^{A(y.x)} + \Delta_{t/0}^{A(x.y)} \tag{25.157}$$

În cazul fenomenelor colective complexe formate sub influenţa a trei factori $A = x \cdot y \cdot z$ aplicarea metodei substituirilor în lanţ pentru determinarea influenţelor factoriale se va putea efectua în trei variante de substituire care vor da rezultate diferite. Schemele generale ale influenţelor factoriale în mărimi relative şi în mărimi absolute se prezintă în schema din figura 25.9.

Metoda substituirilor în lanţ, deşi este una din cele mai controversate metode de separare a influenţei factorilor unui fenomen complex pe bază de indici, se bucură de o largă aplicabilitate practică, mai ales datorită simplităţii ei.

25.6. Indicii teritoriali

Indicii teritoriali măsoară variaţia medie relativă a fenomenelor de acelaşi tip, coexistente în timp, în profil teritorial. Ei au apărut ca urmare a necesităţii comparării în teritoriu a nivelului de dezvoltare economico-socială, în vederea fundamentării deciziilor de repartizare proporţională a forţelor de producţie pe diferite subunităţi teritoriale, de dezvoltare armonioasă, echilibrată şi eficientă a acestora.

În cazul indicilor teritoriali, comparaţia binară se efectuează pentru fenomene de acelaşi tip din punct de vedere esenţial, care se desfăşoară în acelaşi moment sau perioadă de timp şi sunt în spaţii diferite.

Scanned with OKEN Scanner

Problemele metodologice care trebuie rezolvate în vederea construirii indicilor teritoriali sunt asemănătoare celor ale construirii indicilor în general, cu unele particularităţi specifice, rezultate din faptul că nivelurile aceloraşi fenomene situate în spaţii diferite se compară între ele:

- alegerea bazei de comparare;
- alegerea formulei de calcul în funcţie de natura fenomenelor şi indicatorii care se compară;
- alegerea sistemului de ponderare;
- alegerea procedeelor de separare a influenţelor factoriale.

Apariţia şi dezvoltarea fenomenelor în spaţii diferite imprimă acestora unele trăsături specifice legate de condiţiile naturale şi sociale locale, care afectează deci, gradul de comparabilitate a indicatorilor economici şi sociali proveniţi din diferite unităţi teritoriale şi creează dificultăţi sporite în stabilirea criteriilor de alegere a unităţii teritoriale care să constituie bază de raportare. Problema se complică şi mai mult în cazul ţărilor în curs de dezvoltare, în care locul diferitelor unităţi administrativ teritoriale în structura social economică naţională se modifică în mod permanent, datorită avansării pe trepte superioare ale dezvoltării economice, sociale şi politice. Ca urmare, o unitate teritorială administrativă poate servi ca bază de comparaţie o perioadă limitată de timp, care este adesea mai scurtă decât în cazul indicilor dinamicii, ceea ce creează o neconcordanţă între sistemul indicilor teritoriali şi cel al indicilor dinamicii. Alegerea bazei de comparaţie pentru construirea indicilor teritoriali se complică şi mai mult în cazul comparării unor indicatori sociali, economici din ţări diferite care, de regulă, prezintă diferenţe structurale calitative importante. Problema alegerii bazei de comparaţie se simplifică tot mai mult, pe măsură ce se reduce dimensiunea unităţilor teritoriale care se compară şi invers.

În cazul unui fenomen singular x, care se compară între două unităţi teritoriale A şi B, în funcţie de modul de alegere a bazei şi a momentului sau perioadei de timp pentru care se face compararea, se obţine:

$$i_{A/B}^{x_i} = \frac{x_t^A}{x_t^B} \cdot 100 \qquad \text{şi} \qquad i_{B/A}^{x_i} = \frac{x_t^B}{x_t^A} \cdot 100 \qquad (25.158)$$

Indiferent de momentul sau perioada de timp luată în calcul, cei doi indici sunt reversibili:

$$i_{A/B}^{x_i} \cdot i_{B/A}^{x_i} = \frac{x_i^A}{x_i^B} \cdot \frac{x_i^B}{x_i^A} = 1 \qquad (25.159)$$

Scanned with OKEN Scanner

În cazul unui fenomen cantitativ colectiv omogen, în care însumarea fenomenelor simple se poate face direct, indicii teritoriali vor avea următoarele formule:

$$I_{A/B}^{x_{ij}} = \frac{\sum_{i=1}^{n} x_{ij}^{A}}{\sum_{i=1}^{n} x_{ij}^{B}} \cdot 100 \quad \text{şi} \quad I_{B/A}^{x_{ij}} = \frac{\sum_{i=1}^{n} x_{ij}^{B}}{\sum_{i=1}^{n} x_{ij}^{A}} \cdot 100 \qquad (25.160)$$

iar

$$I_{A/B}^{x_{ij}} \cdot I_{B/A}^{x_{ij}} = \frac{\sum_{i=1}^{n} x_{ij}^{A}}{\sum_{i=1}^{n} x_{ij}^{B}} \cdot \frac{\sum_{i=1}^{n} x_{ij}^{B}}{\sum_{i=1}^{n} x_{ij}^{A}} = 1 \qquad (25.161)$$

Deci, teoretic, şi în acest caz se verifică testul reversibilităţii.

În cazul unor fenomene colective, care nu sunt direct însumabile şi pentru care este necesar să se utilizeze un element de comensurare, o pondere y, pentru construirea indicilor teritoriali, se procedează astfel:

$$I_{A/B}^{x(y)} = \frac{\sum_{i=1}^{n} x_{ti}^{A} y_{ti}^{A}}{\sum_{i=1}^{n} x_{ti}^{B} y_{ti}^{A}} \quad \text{sau} \quad I_{A/B}^{x(y)} = \frac{\sum_{i=1}^{n} x_{ti}^{A} y_{ti}^{B}}{\sum_{i=1}^{n} x_{ti}^{B} y_{ti}^{B}} \qquad (25.162)$$

şi

$$I_{B/A}^{x(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{B}}{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{B}} \quad \text{sau} \quad I_{B/A}^{x(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{A}}{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{A}} \qquad (25.163)$$

Se observă că, respectând principiul folosirii aceloraşi ponderi în ambele teritorii comparate, există 4 posibilităţi de construire a indicilor teritoriali.

Dacă indicii teritoriali se utilizează pentru compararea producţiilor unor produse din două ţări, atunci drept ponderi se pot utiliza preţurile unitare exprimate într-o monedă de circulaţie internaţională.

Scanned with OKEN Scanner

Pentru diminuarea influenţei ponderilor asupra valorii indicilor teritoriali ai unor fenomene colective complexe, din punct de vedere teoretic, se pot utiliza procedeele propuse de literatura de specialitate pentru indicii dinamicii.

Mai des utilizaţi în practică sunt indicii teritoriali sub formă de medie geometrică a indicilor teritoriali sub formă agregat, în care se folosesc ponderile celor două teritorii comparate, după formula:

$$I_{A/B}^{x(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{A}}{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{A}} \cdot \frac{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{B}}{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{B}} \qquad (25.164)$$

şi

$$I_{B/A}^{x(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{B}}{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{B}} \cdot \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{A}}{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{A}} \qquad (25.165)$$

În comparaţiile internaţionale, aceste formule se recomandă în cazul unor ţări cu structuri social economice asemănătoare.

Indicii teritoriali se pot calcula şi sub forma unui raport între nivelurile medii aritmetice ponderate ale unui fenomen situat în două unităţi teritoriale diferite:

- indicele bifactorial:

$$I_{A/B}^{\bar{x}(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{A}}{\sum_{j=1}^{n} y_{tj}^{A}} : \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{B}}{\sum_{j=1}^{n} y_{tj}^{B}} \qquad (25.166)$$

şi

$$I_{B/A}^{\bar{x}(y)} = \frac{\sum_{j=1}^{n} x_{tj}^{B} y_{tj}^{B}}{\sum_{j=1}^{n} y_{tj}^{B}} : \frac{\sum_{j=1}^{n} x_{tj}^{A} y_{tj}^{A}}{\sum_{j=1}^{n} y_{tj}^{A}} \qquad (25.167)$$

- indicele unifactorial cu structură fixă:

Scanned with OKEN Scanner

$$I_{A/B}^{\bar{x}(y)} = \frac{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{A}}{\sum_{j=1}^{n} y_{ij}^{A}} : \frac{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{A}}{\sum_{j=1}^{n} y_{ij}^{A}} = \frac{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{A}}{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{A}} \tag{25.168}$$

şi

$$I_{B/A}^{\bar{x}(y)} = \frac{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{B}}{\sum_{j=1}^{n} y_{ij}^{B}} : \frac{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{B}}{\sum_{j=1}^{n} y_{ij}^{B}} = \frac{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{B}}{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{B}} \tag{25.169}$$

- indicele unifactorial cu variaţii ale structurii:

$$I_{A/B}^{y(\bar{x})} = \frac{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{A}}{\sum_{j=1}^{n} y_{ij}^{A}} : \frac{\sum_{j=1}^{n} x_{ij}^{B} y_{ij}^{B}}{\sum_{i=1}^{n} y_{ij}^{B}} \tag{25.170}$$

şi

$$I_{B/A}^{y(\bar{x})} = \frac{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{B}}{\sum_{j=1}^{n} y_{ij}^{B}} : \frac{\sum_{j=1}^{n} x_{ij}^{A} y_{ij}^{A}}{\sum_{j=1}^{n} y_{ij}^{A}} \tag{25.171}$$

25.7. Indicii de plan

Studiul variaţiei fenomenelor în timp se completează în mod necesar cu analiza evoluţiei fenomenelor în raport cu anumite niveluri prestabilite, de regulă prin planuri şi programe de dezvoltare. Indicii construiţi în acest scop poartă denumirea de *indici de plan*. Ca şi cei ai dinamicii, indicii de plan pot fi *indici individuali* şi *indici de grup*.

Scanned with OKEN Scanner

Pentru a construi indici de plan, să considerăm că cele trei momente din evoluţia unui fenomen într-o anumită perioadă se prezintă ca în figura 25.4.

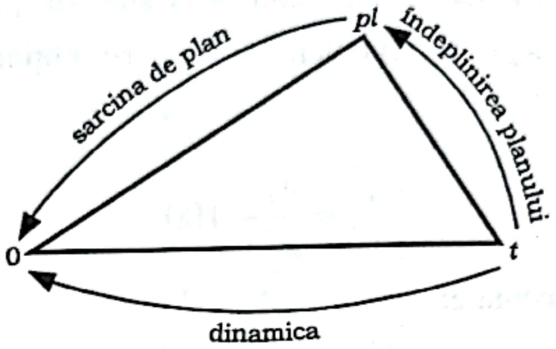


Figura 25.4. Sistemul indicilor de plan

Luând în considerare un anumit fenomen simplu x şi făcând comparaţiile sugerate de săgeţile din figura 25.4 rezultă următorii indici individuali ai planului:

1. Comparând nivelul planificat x_p, al fenomenului x_j, cu nivelul lui din perioada de bază x_0, se obţine *indicele obiectivului sau sarcinii de plan* $i_{p/0}^{x_j}$:

$$i_{p/0}^{x_j} = \frac{x_{pj}}{x_{0j}} \cdot 100$$

În mărimi absolute, sarcina de plan este dată de relaţia:

$$\Delta_{p/0}^{x_j} = x_{pj} - x_{0j}$$

Indicele sarcinii de plan poate fi: $i_{p/0}^{x_j} \gtrless 100$

Dacă: $i_{p/0}^{x_j} = 100$ şi $\Delta_{p/0}^{x_j} = 0$, înseamnă că s-a planificat menţinerea la un nivel constant în perioada de plan a fenomenului x_j;

Dacă: $i_{p/0}^{x_j} > 100$ şi $\Delta_{p/0}^{x_j} > 0$, înseamnă că s-a planificat o creştere a volumului fenomenului x_j faţă de nivelul din perioada de bază;

Dacă: $i_{p/0}^{x_j} < 100$ şi $\Delta_{p/0}^{x_j} < 0$, înseamnă că în perioada de plan s-a prevăzut o reducere a nivelului fenomenului x_j faţă de nivelul existent în perioada de bază.

2. Comparând nivelul efectiv realizat la sfârşitul perioadei planificate x_{ij} cu nivelul planificat x_{pj} al fenomenului x_j, se obţine *indicele îndeplinirii planului* $i_{t/p}^{x_j}$:

$$i_{t/p}^{x_j} = \frac{x_{tj}}{x_{pj}} \cdot 100$$

iar în mărimi absolute:

$$\Delta_{t/p}^{x_j} = x_{tj} - x_{pj}$$

Indicele îndeplinirii planului poate lua valori mai mari, mai mici sau egale cu 100:

$$i_{t/p}^{x_j} \gtrless 100$$

iar $$\Delta_{t/p}^{x_j} \gtrless 100$$

Dacă: $i_{t/p}^{x_j} = 100$ şi $\Delta_{t/p}^{x_j} = 100$, înseamnă că planul a fost îndeplinit întocmai.

Dacă: $i_{t/p}^{x_j} < 100$ şi $\Delta_{t/p}^{x_j} < 100$, înseamnă că planul nu a fost îndeplinit.

Dacă: $i_{t/p}^{x_j} > 100$ şi $\Delta_{t/p}^{x_j} > 100$, înseamnă că planul a fost depăşit.

3. Comparând nivelul fenomenului x_j efectiv realizat în perioada planificată cu nivelul din perioada de bază, se obţine *indicele dinamicii:*

$$i_{t/0}^{x_j} = \frac{x_{tj}}{x_{0j}} \cdot 100$$

iar în mărimi absolute:

$$\Delta_{t/0}^{x_j} = x_{tj} - x_{0j}$$

Scanned with OKEN Scanner

Aceşti trei indici formează un sistem, numit *sistemul indicilor de plan*, legaţi între ei printr-o relaţie bine determinată şi anume, *produsul dintre indicele sarcinii de plan şi indicele îndeplinirii planului este egal cu indicele dinamicii corespunzător:*

$$i^{x_j}_{p/0} \cdot i^{x_j}_{t/p} = i^{x_j}_{t/0}$$

şi

$$\Delta^{x_j}_{p/0} + \Delta^{x_j}_{t/p} = \Delta^{x_j}_{t/0}$$

Într-adevăr:

$$\frac{x_{pj}}{x_{0j}} \cdot \frac{x_{tj}}{x_{pj}} = \frac{x_{tj}}{x_{0j}}$$

şi

$$(x_{pj} - x_{0j}) + (x_{tj} - x_{pj}) = x_{tj} - x_{0j}$$

În cazul unor fenomene complexe, se construiesc indicii de grup ai planului. Ca şi în cazul dinamicii fenomenelor complexe, problema cea mai dificilă este stabilirea ponderilor care urmează a se utiliza în construirea indicilor de grup ai planului. Se pot astfel utiliza ponderi din perioada de bază, ponderi din perioada curentă sau niveluri planificate.

Considerând un fenomen complex $A = x \cdot y$ şi luând în calcul ponderi din perioada de bază, rezultă:

- indicele sarcinii de plan $I^{x}_{A_{p/0}}$:

$$I^{A(x.y)}_{p/0} = \frac{\sum_{j=1}^{n} x_{pj} y_{0j}}{\sum_{j=1}^{n} x_{0j} y_{0j}}$$

şi

$$\Delta^{A(x.y)}_{p/0} = \sum_{j=1}^{n} x_{pj} y_{0j} - \sum_{j=1}^{n} x_{0j} y_{oj}$$

- indicele de grup al îndeplinirii planului $I^{A(x.y)}_{t/pl}$:

$$I_{t/p}^{A(x,y)} = \frac{\sum_{j=1}^{n} x_{tj} y_{0j}}{\sum_{j=1}^{n} x_{pj} y_{0j}}$$

şi

$$\Delta_{t/p}^{A(x,y)} = \sum_{j=1}^{n} x_{tj} y_{0j} - \sum_{j=1}^{n} x_{pj} y_{oj}$$

• indicele de grup al dinamicii $I_{t/0}^{A(x,y)}$:

$$I_{t/0}^{A(x,y)} = \frac{\sum_{j=1}^{n} x_{tj} y_{0j}}{\sum_{j=1}^{n} x_{0j} y_{0j}}$$

iar

$$\Delta_{t/0}^{A(x,y)} = \sum_{j=1}^{n} x_{tj} y_{0j} - \sum_{j=1}^{n} x_{0j} y_{oj}$$

Şi în cazul sistemului indicilor de grup ai planului se verifică o relaţie de acelaşi tip, ca şi în cazul indicilor individuali ai planului: *produsul dintre indicele sarcinii de plan şi indicele îndeplinirii planului este egal cu indicele dinamicii:*

$$\frac{\sum_{j=1}^{n} x_{pj} y_{0j}}{\sum_{j=1}^{n} x_{0j} y_{0j}} \cdot \frac{\sum_{j=1}^{n} x_{tj} y_{0j}}{\sum_{j=1}^{n} x_{pj} y_{0j}} = \frac{\sum_{j=1}^{n} x_{tj} y_{0j}}{\sum_{j=1}^{n} x_{0j} y_{0j}}$$

iar între influenţele exprimate în mărimi absolute:

$$\Delta_{t/0}^{A(x,y)} = \Delta_{A_{p/0}}^{x} + \Delta_{t/p}^{A(x,y)}$$

Aceste relaţii se verifică numai în cazul în care în toţi indicii sistemului se utilizează aceeaşi pondere (y_{tj} sau y_{pj}).

Scanned with OKEN Scanner

Bibliografie

1. AFIFI, A.A., AZEN, S.P. - *Statistical Analysis. A Computer Oriented Approach*, Academic Press Inc., New York, 1979
2. AIVAZIAN, S. - *Etude statistique des dépendences*, Edition MIR, Moscou, 1970
3. ALLEN, R.D.G. - *Index Numbers in Theory and Practice*, The McMillan Press LTD, London, 1975
4. ANDERSON, D.R., SWEENEY, D.J., WILLIAMS, T.A. - *Statistics for Business and Economics*, West Publishing Company, New York, 1987
5. ANDREI, T., STANCU, S. - *Statistica*, Ed. ALL, Bucureşti, 1995
6. ANGER, Pierre - *Tendences actuelles de la recherche scientifique*, UNESCO, Paris, 1961
7. ARDILLY, P. - *Les techniques de sondage*, Editions Technip, Paris, 1994
8. BALLARGEON, G. - *Méthodes statistiques*, vol.I-II, Les Editions SMG, Trois-Rivières, 1985
9. BAGHINSCHI, V. - *Funcţiile de producţie şi aplicaţiile lor în agricultură*, Ed. Ceres, Bucureşti, 1979
10. BAMBERG, G., BAUR, F. - *Statistik*, R. Oldenbourg Verlag, München, 1991
11. BANDETTINI, P. - *I numeri indici*, Edizioni Scientifiche Einaudi, Torino, 1959
12. BARBERI, Benedetto - *Il metodo statistico nello, Studio dei fenomeni osservazionali*, Boringhieri, Torino, 1962.
13. BARBERI, Benedetto - *Elementi di statistica economica*, Edizioni Scientifiche Einaudi, Torino, 1958
14. BĂRBAT, Al. - *Teoria statisticii sociale*, Ed. didactică şi pedagogică, Bucureşti, 1972
15. BARON, T., KORKA, M., PECICAN, E., STĂNESCU, M. - *Statistică*, Ed. didactică şi pedagogică, Bucureşti, 1981
16. BARON, T., ANGHELACHE, C., BARON, T., ANGHELACHE, C., ŢIŢAN, E. - *Statistică*, Ed.Economică, Bucureşti, 1996
17. BARON, T., BIJI, E., TÖVISSI, L., WAGNER, P., ISAIC-MANIU, Al., KORKA, M., POROJAN, D. - *Statistică teoretică şi economică*, Ed. didactică şi pedagogică, Bucureşti, 1996
18. BERTAUD, Marcel - *Introduction à la statistique*, cours programmé 1 et 2, Les Presses de l'Université de Montréal, Montréal, 1973
19. BIALES C. - Analyse statistique des données, Chotard et Associés Ed., Paris, 1987

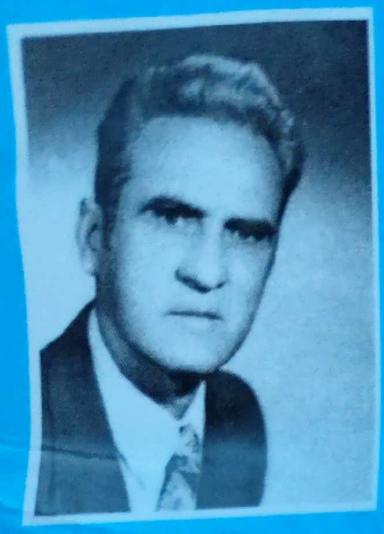
Scanned with OKEN Scanner

46. CROXTON, F.E. - *Elementary Statistics with Applications in Medicine and the Biological Sciences*, Dover Publications Inc., New York, 1953
47. CROXTON, F.E., COWDEN, D.J., KLEIN, S. - *Applied General Statistics*, Prentice - Hall Inc., Englewood Cliffs, N.J., 1967
48. CROW, E.L., DAVIS, F.A., MAXFIELD, M. - *Statistics Manual*, Dover Publications Inc., New York, 1960
49. CYHELSKY, L., KANOKOVA, J. *Teorie statistiky*, Nakladatelstvi technicke literatury, Praha, 1986
50. DAGNELIE, Pierre - *Théorie et méthodes statistiques*, vol. I, II, Ed. J. Duculot S.A., Gembloux, 1969
51. DANIEL, W.W., TERRELL, J. - *Business Statistics*, Noughton Mifflin Company, Boston, 1992
52. DARMOIS George - *Statistique mathematique* , Ed. INSEE. Paris, 1948
53. DAVENPORT, C.B. - *Statistical Methods with Special Reference to Biological Variation*, John Wiley and Sons, New York, 1904
54. DECUSEARĂ, E.C., - *Elemente de statistică teoretică şi aplicată*, Ed. Inst. Central de Statistică, Bucureşti, 1943
55. DEDULESCU, P. - *Noţiuni de statistică generală şi demografică*, Tipografia şi Librăria Românească, Oradea Mare, 1923
56. DRAPER, N.R., SMITH, H. - *Applied Regression Analysis*, John Wiley & Sons Inc., New York, 1966
57. DONDA, A., HERRDE, E., KUHN, O., STRUCK, R. - *Statistik*, Verlag die Wirtschft, Berlin, 1981
58. DUBOS, J. - *Statistique descriptive en sciences économiques*, Dunod, Paris, 1970
59. DUON, Gaston - *De la théorie à la pratique des indices statistiques*, Ed. Eyrolles, Paris, 1955
60. DUSSAIX, A.M., INDJEHAGOPIAN, J.P. - *Méthodes statistiques appliquées à la gestion*, Les éditions d'organisation, Paris, 1981
61. EDWARDS, A.L. - *Statistical Methods*, Holt, Rinehart and Winston, Inc., New York, 1967
62. ELISEEVA I.I., IUZBAŞEV M.M. - *Obşciaia teoria statistiki*, izd. Finansî i Statistika, Moskva, 1996
63. ENACHE, M. - *Sistematizarea teritoriului. Aplicaţii statistice*, Ed. tehnică, Bucureşti, 1986
64. FISHER, Irving - *The Making of Index Numbers*, Houghton Mifflin, Boston, 1922
65. FISHER, Ronald A. - *Statistics for Research Workers*, Oliver & Boyd, Edinborough, London, 1925
66. FLOREA, Ioan - *Statistică*, vol. I, Ed. Universitatea Cluj-Napoca, 1986
67. FOURASTIE, J. - *Les formules d'indices de prix* , Librairie Armand Colin, Paris, 1966
68. FOURASTIE, J., SAGUERY, S. - *Exercices résolus de statistiques*, Masson, Paris, 1993
69. FOURASTIE, J., - *Mathématiques appliquées à l'économie* , Dunod, Paris, 1993
70. FOURASTIE, J. - *Fort en Match* , Delagrave, Paris, 1993

Scanned with OKEN Scanner

149. MUDGETT, Bruce - *Index Numbers*, John Wiley and Sons, Inc., London, 1951
150. MUREŞAN, I. - *Manual de statistică sanitară*, Ed. Medicală, Bucureşti, 1980
151. NAZEM, S. - *Applied Time Series Analysis for Business and Economic Forecasting*, Marcel Dekker Inc., New York, 1988
152. NETER, J., WASSERMAN - *Applied Linear Statistical Models*, Richard D. Irwin Inc., Homewood, Illinois, 1974
153. NICERORO, Alfredo - *La méthode statistique et ses applications aux sciences naturelles, aux sciences sociales et à l'art*, Ed. Marcel Giard, Paris, 1925
154. OLSON, C.L., PICCONI, M.J. - *Statistics for Business Decision Making* - Scott, Foresman and Company, Glenview, Illinois, 1983
155. ONICESCU, O., MIHOC, Gh. - *Lecţii de statistică matematică*, Ed. tehnică, Bucureşti, 1958
156. OPITZ, Otto - *Numerische Taxonomie*, Gustav Fischer vargag, Stuttgart, 1980
157. OSTROM, C.W. - *Times Series Analysis*, SAGE publications, London, 1990
158. PAENSON, Isaac - *Systematic Glossary of the Terminology of Statistical Methods*, Pergamon Press, Oxford, 1985
159. PASQUIER, A. - *Eléments de calcul des probabilités et de théorie, exercices et corrigés*, Dunod, Paris, 1970
160. PASQUIER, A. - *Eléments de calcul des probabilités et de théorie des sondages*, Dunod, Paris, 1969
161. PERROUX, François - *Tehnicile cantitative ale planificării*, Ed. Academiei Române, Bucureşti, 1969
162. PECICAN, E.Ş. - *Econometrie*, Editura ALL, Bucureşti, 1994
163. PIATIER, André - *Statistique*, PUF, Paris, 1962
164. PIATIER, A., CAHUZAC, P. - *Econométrie et mathématiques*, PUF, Paris, 1971
165. PIGANIOL, B. - *Statistique*, Librairie Dalloz, Paris, 1971
166. PLOSKO, B.G., ELISEEVA, I.I. - *Istoria statistiki, Finansî i statistika*, Moskva, 1990
167. POLLAK, R.A. - *The Theory of the Cost-of-Living Index*, Oxford University Press, New York, 1989
168. PRESSAT, Rolland - *Manuel d'analyse de la mortalité* , Ed. INED, Paris, 1985
169. PUPION, Georges - *Statistiques. Annales corrigées*, Dunod, Paris, 1994
170. PURCARU I. - *Matematică şi asigurări*, Ed. Economică, Bucureşti, 1994
171. QUINET, E. - *Séries temporelles et décisions économiques*, Dunod, Paris, 1969
172. RANCU, N., TOVISSI, L. - *Metode statistice de control al calităţii producţiei industriale*, Ed.ştiinţifică, Bucureşti, 1957
173. RANCU, N., TOVISSI, L. - *Statistica matematică cu aplicaţii în producţie*, Ed. Academiei Române, Bucureşti, 1963
174. REIF, F. - *Fizică statistică*, Ed. didactică şi pedagogică, Bucureşti, 1983
175. RESA, Ion - *Statistică*, Ed Universităţii Timişoara, Timişoara, 1986
176. REYNOLDS, H.T. - *Analysis of Nominal Data*, SAGE Publications, Beverly Hills, 1985
177. RICHMOND, Samuel B. - *Statistical Analysis*, The Ronald Press Company, New York, 1964

Scanned with OKEN Scanner



Prof. univ. dr. MIHAI ȚARCĂ

LICENȚIAT în anul 1965 al *Academiei de Studii Economice*, București, Facultatea de *Economie Generală*, Specializarea *Statistică*

DOCTOR ÎN ECONOMIE, Specialitatea *Statistică*, A.S.E. București, 1973

SPECIALIZĂRI ÎN STRĂINĂTATE:

University of Wisconsin, Madison, SUA, 1975; *State University of North Carolina*, Chapel Hill, USA, 1976; Frieburg Universität, Germania, 1984; *University of Nebraska*, Omaha, SUA, 1993; *Université Paris XI*, 1994; *Reading University*, Marea Britanie, 1995; *University of Nijmegen*, Olanda, 1996; *Université Toulouse I*, Franța, 1997.

ALTE LUCRĂRI PUBLICATE DE ACELAȘI AUTOR:

1. *Statistică teoretică și economică. Culegere de probleme*, Iași, 1970;
2. *Introducere în prognoza demografică*, Iași, 1974;
3. *Universitatea „Al.I. Cuza" Iași*, Iași, 1977, autor și coordonator;
4. *Statistică*, vol. I și vol. II, Iași, 1979;
5. *Satistică*, București, 1980, în colaborare;
6. *The „Al.I. Cuza" University of Iași*, Iași, 1981, autor și coordonator;
7. *Privatisation and the Market Economy*, Iași, 1993, autor și coordonator;
8. *Romanian Population, Past, Present and Future Trends*, Iași, 1993;
9. *Regression et correlation*, Paris, 1994;
10. *Demografie*, București, 1997.

Tratatul de Statistică Aplicată, pe care îl punem cu acest prilej la dispoziția cititorilor, publicat într-o primă variantă în urmă cu aproape două decenii, este conceput ca o lucrare de sinteză a noțiunilor, principiilor și metodelor statistice, oferind astfel o imagine cuprinzătoare atât asupra stadiului atins de această știință în prezent, cât și asupra posibilităților și limitelor în care pot fi folosite metodele statistice pentru analiza și prognoza dezvoltării fenomenelor social-economice.

Fiind o lucrare cu caracter metodologic, *Tratatul de statistică aplicată* îmbină în mod armonios prezentarea noțiunilor teoretice cu necesitățile practicii economice și sociale. Expunerea teoretică este completată cu aplicații practice bazate pe date reale din statistica românească și din cea internațională, iar rezultatele sunt analizate și uneori ilustrate grafic. Acest mod de tratare sperăm să fie de un real folos cititorilor din domenii variate ale științelor sociale ca și celor care studiază fenomenele naturii.

Având în vedere această orientare, lucrarea poate constitui un îndreptar prețios și o continuare în direcția aplicațiilor practice atât pentru lucrările teoretice de statistică matematică, informatică și econometrie, cât și pentru lucrările aplicative de economie, demografie, antropologie, sociologie, biologie etc. în vederea cuantificării, prelucrării științifice, analizei și prognozei fenomenelor din natură și societate pe baza informațiilor existente în aceste domenii.

Lei 70000

ISBN 973-30-5275-2

Scanned with OKEN Scanner